Review Machine learning in photosynthesis: prospects on sustainable crop development

Ressin Varghese[1], Aswani Kumar Cherukuri[2], Nicholas H Doddrell[3], C. George Priya Doss[1], Andrew J. Simkin[3,4], Siva Ramamoorthy[1]✉

[1] School of Bio Sciences and Technology, VIT University, Vellore 632014, Tamil Nadu, India

[2] School of Information Technology and Engineering, VIT University, Vellore 632014, Tamil Nadu, India

[3] School of Biosciences, University of Kent, Canterbury, United Kingdom, CT2 7NJ, UK

[4] School of Life Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

✉ Correspondence addressed to: siva.ramamoorthy@gmail.com (Siva Ramamoorthy)

Improving photosynthesis is a promising avenue to increase food security. Studying photosynthetic traits with the aim to improve efficiency has been one of many strategies to increase crop yield but analyzing large data sets presents an ongoing challenge. Machine learning (ML) represents a ubiquitous tool that can provide a more elaborate data analysis. Here we review the application of ML in various domains of photosynthetic research, as well as in photosynthetic pigment studies. We highlight how correlating hyperspectral data with photosynthetic parameters to improve crop yield could be achieved through various ML algorithms. We also propose strategies to employ ML in promoting photosynthetic pigment research for furthering crop yield.

Keywords:

Photosynthesis; Machine learning; Crop yield, Deep learning; Photosynthetic pigments

Abbreviations

ANN- Artificial neural network, CNN, FCN- fully convolutional neural network, DGCNN-Dynamic Graph, ELM- extreme learning machine, CP-ANNs- counter-propagation artificial neural networks, EN- Ensemble learning, GBRT- gradient boosting regression tree, GPR- Gaussian process regression, KELM- kernel-based extreme learning machine, k-NN- k-nearest neighbors, ML- Machine learning, MLP- Multilayer Perceptron, MODIS-Moderate-resolution imaging spectroradiometer, NBC- Naive Bayes Classifier, MLR-Multiple linear regression, NDVI- normalized difference vegetation index, NEE- Net ecosystem exchange of CO_2, PCA- Principle component analysis, PLSR- partial least square regression, QTL- Quantitative trait locus, RBFN- Radial Basis Function Networks, RF- Random Forest, ENET- elastic net, RFR- random forest regression, RGB- Red, green, blue, SGB-Stochastic gradient boosting, SKNs- Supervised Kohonen Networks, SVM- support vector machines , SVR- support vector regression, TE- Transposable elements., UAV- Unmanned aerial service, XY-Fs- XY-fused Networks.

1. Introduction

Global food production needs to be increased to feed the growing population by overcoming fluctuating climate changes, decreasing yield productivity, low feed stocks, small rural labor forces, depleted soil fertility, a loss of available agricultural farmland to other uses, reduction in available water resources, and reduced efficacy of agrochemicals [http://www.fao.org › wsfs, Zhang et al., 2021). It is obligatory to meet agricultural demands by augmenting crop yield on a global scale. The important role of photosynthesis and photosynthetic pigment research in increasing crop yield has been at the center of many studies (Long et al., 2006). Remarkable outcomes have been reported by modifying the Calvin Benson Cycle (Simkin et al., 2015; López-Calcagno et al., 2020) photorespiration (Simkin et al., 2017b; Lopez-Calcagno et al., 2018), and increasing photosynthetic electron transport rates (López-Calcagno et al., 2020; López-Calcagno et al., 2018). Strategies to augment photosynthetic rate and biomass with particular reference to photosynthetic pigments have also been investigated in many studies (Simkin et al., 2022). For instance, rice (*Oryza sativa*) mutants expressing decreased chlorophyll levels formed chloroplast with elevated gene expression of thylakoid membrane proteins. Higher levels of these proteins, which are involved in chlorophyll-binding, lead to increased photosynthetic rate and better canopy light distribution.

The advent of numerous micrometeorological techniques followed by optical, reflectance sensors aided in the determination of photosynthetic activity. Many of these studies have generated large data sets and subsequently, however meaningful analysis is complicated and time-consuming. Machine learning (ML), one of the most discussed technical advancement of the century have aided in predicting photosynthetic activity from plant cells to large terrestrial ecosystems by interpreting and analyzing these huge set of data. However, there are no comprehensive reports available highlighting the various domains of ML-driven photosynthesis and photosynthetic pigment research. Here we review how ML-oriented approaches can facilitate photosynthesis research and in turn, accelerate improvements to crop yield and biomass. We also propose possible ways of using ML algorithms to improve photosynthetic pigment research.

2. How do photosynthesis and photosynthetic pigments improve crop yield?

Maximum yield potential is achieved when optimum conditions are provided to the crop in the absence of stresses (Gu et al., 2017). Long et al (Long et al., 2006) determined the maximum yield potential by identifying and quantifying the photosynthetic parameters that determine crop yield and defined this as the ‘yield equation’. The yield equation is defined as follows:

$$Pn = St*\varepsilon i*\varepsilon c/K$$

$$Yp = \eta*Pn$$

where Yp (yield potential), η describes the harvest index (i.e. biomass partitioned into the harvestable material), Pn (primary production of biomass), St (incident solar radiation over a crop), εi (efficiency of light interception by the crop determined by the photosynthetic light absorption characteristics of the leaves like leaf area), εc (efficiency of conversion of intercepted light into biomass which is dependent on the wavelength of light absorbed i.e. pigment types and ratios) and K (energy content of the harvestable biomass). Many of these parameters, including harvest index and interception efficiency, are nearing their theoretical maximum (Morgan 2005; Dermondy et al., 2008; Zhu et al., 2010). However, εc was found to be ~30 % of its theoretical maximum.

Considerable studies have proven that increasing CO_2 assimilation, inhibiting oxygenase activity of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) to reduce photorespiration, hasten carboxylation, reduce stomatal conductance could improve photosynthesis and thus, ultimately, crop yield (South et al., 2018; Simkin et al., 2019; Weber et al., 2019; Raines et al., 2022). Higher photosynthetic rates correspond to higher assimilation of atmospheric [CO_2], providing additional carbon for the production of secondary metabolites required for growth. Work has previously established, that increasing CO_2 uptake, for example through the growth of crops in elevated [CO_2] (Doddrell et al., 2023; Mortensen et al., 1994; Dong et al., 2020; Ainsworth and Long, 2005) or through genetic engineering (Simkin et al., 2019; Simkin, 2019; Raines, 2023), increases biomass and crop yield. For example, genetic engineering studies in the Calvin-Benson cycle targeting Sedoheptulose-1,7-bisphosphatase (SBPase) have corroborated the role of Photosynthesis in crop yield augmentation (Simkin et al., 2017b; Lefebvre et al., 2005; Driever et al., 2017). Of keynote was the 53% increase in seed yield (Simkin et al., 2017b) and 40% increases in grain yield (Driever et al., 2017; Simkin et al., 2019) observed in Arabidopsis and wheat over-expressing SBPase respectively, demonstrating the significant increase in yield that can be achieved through genetically manipulating photosynthetic carbon uptake. Thus, improving photosynthesis is the most consistent way to improve crop productivity (Simkin et al., 2019; Simkin, 2019).

The electron transport system is another potential target that could enhance carbon assimilation, photosynthesis, and crop yield. For example, genetic manipulations of the Cytochrome b_6f (cyt b_6f) complex by over-expression of the Rieske iron-sulfur protein has been reported to be an efficient approach to increasing biomass yield and seed yield (Simkin et al., 2017a; Simkin et al., 2019). Furthermore, the expression of the algal cytochrome c_6, which functions as an electron carrier between cyt b_6f and the PSI reaction center, in Arabidopsis (Chida et al., 2007) and tobacco (Yadav et al., 2018; López -Calcagno et al., 2020) has been shown to increase the rate of plant development and increase biomass yield in greenhouse and field experiments.

Moreover, photosynthetic pigments such as chlorophyll, carotenoids, and phycobilin play a significant role in light harvesting, Photosynthesis, and productivity (Morgan et al., 2005; Simkin et al., 2022). The different absorption spectra of these pigments open the possibility of engineering plants according to environmental conditions and specific light availability. The properties of pigments, such as preventing photodamage and shortening photoinhibition to dissipate excess energy, make them a suitable tool to enhance photosynthetic efficiency (Simkin et al., 2022).

Chlorophyll (*Chl*) is an essential photosynthetic pigment and largely determines photosynthetic capacity and secondary photosynthetic pigments, carotenoids (β-carotene, zeaxanthin, violaxanthin, and lutein) contribute to both light harvesting and photoprotection (see Simkin et al 2022 for review). These pigments have previously been shown to directly influence yield and act to improve yield by three mechanisms. Firstly, Chlorophyll a (*Chl a*) and Chlorophyll b (*Chl b*) absorb sunlight at different wavelengths (*Chl a* absorbs red-orange light; *Chl b* absorbs blue-purple light), suggesting that the total amount of chlorophyll in the leaves (*Chl a* + *Chl b*) directly influence the photosynthetic capacity (Li et al., 2018; Croft et al., 2017). Furthermore, the allocated ratio (*Chl a/b*) directly influences light capture at specific wavelengths. For example, the downregulation of the enzyme chlorophyllide an oxygenase (CAO), responsible for *Chl a* to *Chl b* conversion in plants, reduced the accumulation of *Chl b* increasing the ratio of *Chl a/b* and reducing the light-harvesting antenna size (Ayumi Tanaka, 1998; Simkin et al., 2022). Plants with a smaller antenna size outperformed wild-type plants achieving a 40% increase in biomass yield (Friedland et al. 2019). Secondly, carotenoids harvest violet and blue-green light (400−550 nm) (Hashimoto et al., 2016) and transfer the energy to *Chl a*, increasing the spectrum of light absorbed by the light-harvesting complex (Domonkos et al., 2013). Finally, zeaxanthin, violaxanthin, and lutein play a role in quenching excess energy (Non-photochemical protection; NPQ) and protecting the photosystems from oxidative damage often ascribed to the formation of reactive oxygen species (ROS) in dynamic fluctuating environments (Niyogi et al., 2004; Simkin et al., 2022; Krieger-Liszkay et al., 2008; Hashimoto et al. 2016; Ledford and Niyogi 2005). Speeding up plant recovery to fluctuating light in field conditions resulted in a 20% increase in biomass yield in tobacco (Kromdijk et al., 2016; De Souza et al., 2022).

Photosynthesis improvement studies primarily focus on the rate of carboxylation (Vc_{max}), the electron transport rate (J_{max}), and the mesophyll conductance for carbon dioxide [14]. However, augmenting these aspects in the field and resultant outcomes depend on multiple factors such as temperature, water availability, pathogen attacks, and nutrient content for example. Coupling these data together, creating ML models, and translating them to the global level, can be achieved, to a greater extent by sophisticated ML algorithms.

3. Current status of ML in plant science

The world witnessed the rising of artificial intelligence (AI) during the mid-1950s, and it has rapidly been assimilated into all walks of life. The machines that perform calculations were the first breakthrough in the long-term human-machine alliance, and they adopted different forms becoming an integral part of our lives (Naqa & Murphy, 2015). AI is intended to mimic human intellectual behavior using machines for automating the tedious processes in our daily life. However, AI now surpasses human capabilities and is one of the widely used technologies for data analysis in various science domains. Machine learning algorithms regarded as the new facet of artificial intelligence enable machines to handle and interpret complex data. Robust computational algorithms in ML supersede human brains in predicting outcomes of input data in a very meticulous and well-defined manner. ML can also be defined as the improved form of old-school statistics and regression models. **Box 1** explains various types of machine learning approaches that are currently being used. Interestingly, research in plant science has been drastically improved by the introduction of ML in the domains of pathology, phenology, species detection, herbarium studies, and plant genomics, for example (Dey, 2016).

Machine learning-aided plant science studies have rapidly advanced with the introduction of remote sensing technologies. This aspect of agritech has now revolutionized large-scale studies and is reflected in computer-aided models for combining grasslands and cropping models, livestock models, pest and disease models, and stress phenotyping models aimed at improved productivity (Gonalez- Camacho et al., 2018). For example, the prognosticative nature of ML has been well explored in detecting weeds, disease, and pathogen attacks and in identifying superior mutants in plant breeding studies (Liakos et al., 2018). In genomics, ML is a good prediction tool to determine the functions and regulation of plant genes. For instance, in studying cis-regulatory elements, the activity of gene promoters has been defined (Uygun et al., 2019). Table 1. summarises some of the major applications of ML in plant science.

4. ML: Interlink between plant optical properties and photosynthetic activity

Plant optical measurements owing to the chlorophyll content are critical in the determination of photosynthetic activity. Spatial, spectral, and temporal optical estimations provide us with an outlook of photosynthetic activity. A wide array of spectroradiometers ranging from handheld spectroradiometers to airborne and spaceborne hyperspectral cameras are currently available. The two peaks of chlorophyll fluorescence (near 685 and 740nm) varying with the photochemical reactions within a chloroplast is the parameter assessed through these instruments (Raychaudhiri, 2012; Khurschev et al., 2022). Multispectral spectroradiometers installed in satellites are the most prominent among them. For instance, the Moderate-resolution imaging spectroradiometer (MODIS) used on TERRA as well as in NASA satellites possesses 36 spectral bands of 400 to 14,385 nm with spatial resolution from 250 to 1000 m (Justice et al., 1998). Air bone spectral imaging programs like the National Ecological Observatory Network Airborne Observation Platform with the spectrometer of range 380 to 2500nm measuring 426 bands are also available (Kampe et al., 2010). Satellite vegetation fluorescence and absorption measurements are extremely obliging in assessing the photosynthetic activity of large terrestrial areas. We encourage readers to refer more about spectroradiometers and remote sensing in the review Fu et al., 2022, Siebers et al., 2021.

Hyperspectral reflectance is the most utilized tool to evaluate photosynthesis activities due to its non-destructive nature. There are significant concerns about utilizing these data in canopy scales (Schlund et al., 2020). The spectral reflectance at canopy levels is influenced by numerous factors including plant geometry, architecture, leaf nature, and soil resulting in spurious spectral variations and blurring of the spectral signals. Besides, sensor-based photosynthetic measurements are now centered on data mining of spectral information. Various statistical models founded on ML algorithms correlating these hyperspectral data with photosynthetic activity came into use for effortless data elucidation (Siebers et al., 2021; Fu et al., 2022). Partial least square regression models exploring the rapport between dependent and independent variables using latent variables are the most utilized among them. The utilization of various ML algorithms in photosynthetic research has been discussed in the subsections below. An overview of ML in photosynthetic research has been depicted in Fig. 1. Table 2. lists the various plant varieties and ML models utilized for photosynthesis estimation.

4.1 ML in predicting global photosynthesis

The advent of high throughput remote sensing techniques has widened the exploration of photosynthetic machinery from small plant cells to a global scale. However, it is a prerequisite to mention that the gas exchange measurements estimating the CO_2 exchange were the preliminary parameter utilized before the arrival of remote sensing techniques. There are manual and auto chamber methods to evaluate the CO_2 flux in limited spatial areas. In addition, Eddy covariance is the prominent micrometeorological approach applied to estimate the Net ecosystem exchange of CO_2 (NEE) (Siebers et al., 2021) in large areas. It measures the CO_2 flow of rotating eddies over canopies producing the flux footprint of that area. NEE is defined as the difference between CO_2 released by all respiration processes (RECO) and carbon used by Photosynthesis (GPP). While GPP is the principal flux of earthly carbon uptake and therefore understanding global energy fluxes advances photosynthetic research (Schlund et al., 2020). NEE is a highly considered factor in global energy flux and photosynthesis studies. ML is an established tool to determine NEE at a global scale because of its ease in identifying key variables for predicting NEE, combining field measurements with satellite data avoiding observational constraints. Random Forest (RF) models have been used to study the temporal variation of NEE in the Great Basin region. It was understood that downward solar radiation, leaf area index, and soil moisture are important variables for predicting NEE (Zhou et al., 2019). Likewise, Tramontana et al. have used an ANN-based approach to analyze the partitioning of NEE into RECO and GPP in FLUXNETs. Artificial Neural Network (ANN) was suggested as a better ML method to estimate GPP, RECO without any assumption of driver-output relationships using soil and meteorological variables (Tramontana et al., 2020).

ML-based GPP models have also been incorporated to investigate FLUXNET (CO_2 flux networks at various ecosystems) in the name FLUXCOM (Tramontana et al., 2016). FLUXCOM targets to evaluate the uncertainties in empirical upscaling and to develop an ensemble of ML-based Flux products. Spectral data from Moderate Resolution Imaging Spectroradiometer (MODIS), and eddy- covariance data, were utilized and some of the major variables considered were Daytime land Surface Temperature and photosynthetic active radiation in the study. Additionally, ML regression algorithms such as neural networks, regression splines, and tree algorithms have been well utilized in estimating carbon and energy fluxes (Xie et al., 2020).

As aforementioned, hyperspectral analysis is now widely utilized to access the photosynthetic performance of interleaf, plant, canopy, and an entire ecosystem. The photosynthetic rate of tomatoes has been predicted in a greenhouse environment from the parameters of growth temperature, available humidity, photon flux density, and CO_2 content using the SOPSO-LSSVM algorithm on a small scale (Liu et al., 2021). However, estimating global photosynthesis is an intricate and transdisciplinary task applying remote sensing, biochemistry, and plant physiology. Various gross primary production (GPP) predictive models such as process-based models (PBMs), semi-empirical light use efficiency (LUE) models, data-driven statistical models, and models employing spectral data have been used to assess global photosynthetic activity.

Several satellite datasets utilizing solar-induced chlorophyll fluorescence (SiF), a fraction of absorbed photosynthetic active radiation (FAPAR) incorporated with the ML approach helped in determining GPP, photosynthesis rate, and global biomass to a great extent (Forkel et al., 2019). The combined approach involving SiF and ML is considered a benchmark in the history of global photosynthesis prediction models. For instance, latent and sensible heat flux along with GPP was estimated using ANN and SiF. ANN was trained using flux and GPP estimated from 2008-2010. The input datasets utilized monthly prediction on a global scale were majorly SiF, net radiation, soil moisture, the temperature of the air, and precipitation (Alemohammad et al., 2017).

Seasonal fluctuations are reported to have a considerable impact on photosynthesis in diverse ecosystems. Particularly late season photosynthesis or ending date of Photosynthesis (EOP) is largely affected by temperature and water variations. Recently, ML-aided reconstructed contiguous SIF data was used as a dataset as a proxy of GPP to evaluate the correlation between temperature, water restraints, and EOP. SIF Data from the Orbiting Carbon Observatory 2 (OCO-2) level 2 SIF and collocated reflectance data from the Moderate Resolution Imaging Spectroradiometer (MODIS) from 2015 to 2016 were used as the training data set for the studies employing Support Vector Machine algorithms (SVM). The different datasets used for the study were CSIF dataset generated using the neural network from four bands reflectance from MODIS for 2001 to 2017, Climate datasets from ERA-Interim reanalysis data (Type 1) and remote sensing based (Type 2), FLUXNET2015 date set from daily GPP estimates through night time partitioning method and the reference Ustar. They also utilized smoothing and threshold-based methods to calculate EOPs from both CSIF and EC-based GPP estimates. Finally, SVM was used to predict whether the EOP or late growing-season photosynthesis is limited by water or temperature after calculating correlations between pre-EOP climate and EOP. The results showed that late-season photosynthesis is significantly influenced by water availability suggesting the need for soil water for improved vegetation during late-season photosynthesis (Zhang et al., 2020).

Besides, ML-based GPP models like MTE (Model Tree Ensembles), RF, ANN, and Multivariate Adaptive Regression Splines (MARS) using time series Standardized Precipitation–Evapotranspiration Index data were utilized to study the lagged effect of drought on photosynthesis in different regions. The studies showed that the tropical region (20°N–20°S) has lower reliabilities of lagged effect regions, highlighting the need for accuracy in assessing effects (Xie et al., 2020). It has also been evidenced that the Gradient Boosted Regression Tree (GBRT) algorithm is efficient in reducing uncertainties in rescaling, and observation-based data in predicting GPP (Schlund et al., 2020). Interestingly, neural networks have used Sentinel-2 and Landsat 8 satellite data to study and develop GPP predictive models to estimate crop primary productivity (Wolanin et al., 2019).

Undeniably, all the above studies have underlined the fact that ML-based models are a better tool to predict energy fluxes, GPP, and biomass distribution at the global level and thus assure more plant productivity. In a nutshell, ML-based photosynthesis prediction models are an asset to plant physiology, precision agriculture studies, and in estimating global GPP. Fig. 2 illustrates the use of ML in predicting photosynthesis.

4.2 Using ML to improve plant physiology and productivity

The photosynthetic capacity of plants is considered a criterion for selecting superior breeding lines and wild cultivars for crop cultivation. As mentioned earlier, rubisco carboxylation and maximum electron transport rate are determinative factors to characterize photosynthetic efficacy in C3 plants. Conventional approaches to measuring these factors using leaf gas exchange are now enhanced via hyperspectral sensor-aided high-throughput phenotyping methodologies (HTP). Leaf reflectance, leaf phenotypic traits, and SIF are utilized to evaluate photosynthetic efficacy by remote sensing techniques (Fu et al., 2021). The ML-enabled photosynthesis research approach has considerably eased the arduous task of interpreting these spectral data and resolved the phenotyping bottleneck to a large extent. HTP is also considered an efficient tool to evaluate crop growth, and physiological changes on par with seasonal variations (Fu et al., 2022). Interestingly, hyperspectral data coupled with ML algorithms are also applied to predict genetic variations and to select plants with improved photosynthetic traits (Silva-Perez et al., 2018; Furbank et al., 2021).

Combined spectroscopic techniques and regression analysis were extensively utilized to screen germplasms with increased photosynthetic potential (Ainsworth et al., 2014; Yendrek et al., 2017). PLSR models using linear multivariate models are applied mostly for predicting photosynthetic parameters because of their precision in dealing with irrelevant spectral bands, band collinearity, and higher R^2 values. Leaf reflectance from *Brassica oleracea* (C_3) and *Zea mays* (C_4) was employed to evaluate the photosynthetic capacity using partial least square (PLS) regression models. In the above study, Intraspecies photosynthetic capacity models predicting carbon-nitrogen ratio, leaf water, and leaf rubisco level were successfully implemented, underlining the significant role of ML algorithms in photosynthesis prediction models (Heckmann et al., 2017). Investigation in tobacco varieties has proved that PLSR with inputs of spectral indices was efficient in distinguishing the photosynthetic capabilities of plants when compared to PLS with reflectance spectra and numerical model inversions (Fu et al., 2020). PLS models are very effective in predicting Vc_{max} in *in-vivo*, *in-vitro*, wild, and in genetically modified plants. Chlorophyll, carbon, and nitrogen content can also be evaluated by building PLS models independently for different spectral regions (Sexton et al., 2021). Meacham-Hensold et al. reported the use of the PLSR model in genetically engineered tobacco varieties (Rubisco antisense varieties) to predict Vc_{max}, $J_{max,}$ and interannual photosynthetic variations. The training data set was built in 2016 from leaf spectral reflectance, gas exchange, and nitrogen content (Feret et al., 2019). The study suggested the appropriateness of the application of PLSR models to more crop varieties for studying the photosynthetic variations in farming landscapes.

Later, various ML algorithms were more utilized considering the pitfalls in PLSR. The support vector machine (SVM) algorithms have been used to study leaf mass per area and equivalent water thickness for analyzing plant functioning, and ecosystem processes at the canopy level using reflectance and transmittance data (900-2400nm) as an alternative to PLSR (Féret et al., 2019). The pitfalls in PLSR have been investigated by substituting stacked regression algorithms such as artificial neural network (ANN), most minor absolute shrinkage and selection operator (LASSO), Gaussian process (GP), SVM, and RF. This regression stacking was applied to tobacco cultivars for studying photosynthetic parameters and it was found that R^2 value and predication potential were increased (Fu et al., 2019). It is suggested that a combined approach of ML algorithms must be applied to optimize the application of ML for improving photosynthesis and deriving more favorable outcomes for photosynthetic research.

Forecasting crop yields based on agriculture expert systems, and machine learning algorithms is a promising approach in precision agriculture. Interestingly, photosynthetically active radiation (PAR) is a well-studied parameter to analyze the leaf area index, which denotes crop growth. Regression trees and ANN are precise, robust ML tools, reported as being efficient in leaf area modeling of crop canopies and forecasting crop yield through analyzing PAR data (Dunea & Moise, 2008). Photorespiration is another physiological phenomenon undergoing extensive studies to augment crop yield (Walker et al., 2016). A multifactorial ML-based approach to study photorespiration was tested with cucumber and the XGBoost algorithm was identified as a better choice using parameters such as CO_2 concentration, temperature, and leaf position (Zheng et al., 2021).

Nowadays, plant phenomics is another aspect of crop improvement program enabling crop physiologists and breeders to develop better cultivars. Radiation use efficiency, and photosynthesis capacity are some factors determining plant phenomics. High throughput phenotyping along with ML algorithms is now widely utilized in plant phenomics studies. For instance, ML-based algorithms have been used to estimate photosynthetic efficiency in sorghum and wheat using the rate of stomatal conductance and canopy reflectance. Stomatal conductance g_s (the rate of carbon dioxide going in and water vapor coming out through stomata), canopy temperature and yield rate share a direct proportionality showing the possibility of using *gs* as a parameter for more photosynthesis-related crop improvement studies (Furbank et al., 2019). Fig. 3 demonstrates the various datapoints utilized to generate ML models to study plant physiology and productivity

4.3 Studying photosynthetic capacity under stress

Photosynthetic capacity under stress conditions is an intense domain of research because the adverse impacts of extreme climatic conditions experienced globally can significantly impact yield and therefore food security. With current trends in global climate, an estimated increase in global temperatures of 2°C by 2050, an expected concomitant increase in atmospheric [CO_2] to 550 ppm (IPCC, 2007; Quéré et al., 2009), and an increase in the frequency and severity of droughts and heat waves (Field et al., 2014) over the same period, we will see an increase in periods of intense stresses brought to bear on crop growing regions. Meta-analysis has previously shown a trend that wheat productivity and grain yield decreased with increasing temperature (Asseng et al., 2015).

ML-integrated spectral and gas exchange studies are less troublesome approaches to investigating the impact of biotic and abiotic stress on crops. The photosynthetic CO_2 assimilation rate, stomatal conductance, and transpiration rate of citrus plants were estimated using the RF algorithm and other ML algorithms to establish a photosynthetic predictive model to appraise the drought stress response of plants (Zhou et al., 2021). Another study measured Vc_{max}, and J_{max} to assess photosynthesis by using four ML regression models PLS, BR, LASSO, and ARDR. Reflectance spectra from high night temperatures, CO_2 exposed peanuts and drought stressed sorghum were employed (Buchaillot et al., 2022). Interestingly, waterlogging resistance of poplars was assessed utilizing photosynthetic parameters and SVM, LASSO models (Xie & Shen, 2021). Photosynthetic efficiency and the accompanying growth rate of the plant also depend on the availability of macro and micro-minerals. It is worthy of mentioning the use of ML in mineral element estimation, particularly nitrogen (a parameter of photosynthetic potential) (Chlingaryan et al., 2018). RF algorithms have been applied in determining the nutrient composition of valencia orange leaves by analyzing spectral data at 380-1020 nm (Osco et al., 2020).

The variation in chlorophyll content under stress is an indispensable parameter for analyzing growth rate, and vegetation physiology and identifying deficiencies determining the need for additional nutrients for plant growth. For instance, the correlation and chlorophyll content and drought stress were studied in winter wheat varieties utilizing 9 ML models such as Ridge regression with cross-validation (RidgeCV), Ridge, Adaboost Regression, Bagging Regressor, GBR, RF, SVM, LASSO, and K-Neighbor. Besides, SVM was the best model under water-limited conditions and Ridge CV showed higher precision under normal irrigation (Wang et al., 2022). A collective of spectral (RGB, fluorescence, thermal, hyperspectral) (Zubler &Yoon, 2020) and ML algorithms have significantly eased photosynthetic pigment composition studies. Large-scale sensing methodologies have been employed to assess chlorophyll content in large rice fields (An et al., 2020). The chlorophyll concentration in rice fields was studied using ML algorithms including GPR (Gaussian process regression), RFR (random forest regression), GBRT (gradient boosting regression tree), and SVR (support vector regression), and optimization of algorithms was through the training data set and grid search and cross-validation. Moreover, the above study suggested that RFR and GPR are more effective algorithms for exploring chlorophyll content under field conditions (An et al., 2020). Similar studies were reported in tea leaves where RF, SVM, deep belief nets, and KELM (kernel extreme learning machine) were compared; KELM (using MATLAB and Statistics Toolbox), integrated with hyperspectral reflectance, was proved to be the proficient strategy to detect chlorophyll content (Sonobe et al., 2020). Chlorophyll concentration has also been estimated analogously in other plants like sorghum using PLSR, RFR, SVR, and ELR which emphasized the efficiency of ML-based spectral studies in pigment estimation (Bhadra et al., 2020). These studies provide a non-destructive means of efficiently estimating leaf chlorophyll concentration, plant development, vigor, and nutrient requirements in the field that will provide the necessary data to breeders and scientists to, one, improve growing and fertilization practices, two, improve plant health and consequently plant yield, and three, apply these advancements to other plants and studies. Fig. 4 shows the utilization of ML in studying photosynthetic capacity under stress.

4.4 How to use ML for augmenting crop productivity through photosynthesis and photosynthetic pigment research?- An insight

What will be the status of the agricultural sector in 2050? The successful implementation of machine learning in photosynthetic research could ignite another green revolution that will significantly increase data collection, analysis, and photosynthetic mapping on a global scale and thus provide potential solutions to elevate crop yield. An approach using stacked ML algorithms will produce more effective, robust results. Extensive investigations are recommended for studying the photosynthetic capacity variations in different seasons using a combined ML, hyperspectral reflectance data approach.

Photosynthesis is a cascade of reactions involving many enzymes, and organic, and inorganic molecules. Amin et al. (2022), utilized decision Tree Classifier and K-means clustering models to predict the oxidation states of Mn ions in the oxygen-evolving complex of photosystem II to understand water splitting reaction. It was predicted that Mn1 and Mn4 are more likely to be oxidized during the transitions from S1 to S2 and S2 to S3 states (Amin, 2022). Promising research to understand photosynthetic machinery utilizing ML techniques will enable the creation of artificial photosynthesis systems. The hyperspectral data combined with training data sets derived from pulse amplitude-modulated chlorophyll fluorescence (PAM) system via DL could be used to study the genetic diversity of plants like wheat, to identify traits that could be introduced into new varieties via breeding programs, genetic engineering or gene editing, to enhance productivity and nutritional quality. ML must be explored extensively to study the QTLs and SNPs and thus the genetic framework of Photosynthesis. The efficient implementation of ML algorithms could ease the burden of countless *in-vivo, in-vitro* trial, and error experiments to market a new variety. The intricate correlations between photosynthetic traits, N portioning, flowering variation, and phenology (Furbank et al., 2020) should be subjected to DL-like algorithms to derive a timely relationship. ML is the key to precision agriculture, and thus we could envisage a world where the farmer can detect the phenotypic and genotypic variations in the field in a far more exacting level of detail than is currently possible.

The momentum in photosynthetic research ignited by ML could- be extended to photosynthetic chlorophylls and accessory carotenoid, xanthophyll pigments. Unfortunately, the ML approaches in pigment studies are still in the early stages. ML could magnify our requirements to predict, identify, and classify stress in combination with high throughput automated sensor techniques with particular reference to photosynthetic pigments. Currently, advanced images are deployed for stress phenotyping by counting leaf lesions, as an estimation of the severity of the stress at the leaf surfaces (Singh et al., 2021). The variations in pigments on exposure to stress factors could be used as a parameter to study the severity of stress on plants. Implementation of DL algorithms in interrogating stress-pigment association will incontrovertibly amplify pigment biology and crop yield. The prognostication aspect of ML could be exploited aptly in picturing the possibilities of pest attacks and diseases by employing pigment content as a parameter (Aparecido et al., 2020)

Moreover, a predictive model based on pigment fluxes in the incidence of pest attacks and diseases will be ground-breaking. The invasion of weed species in large farm areas of crop plants could be detected earlier using ML (Baron et al., 2018). Nutritional imbalances in plant growth could also be assessed through neuro-fuzzy logic ML algorithms (Garcia-Perez et al., 2020). Last, but not least, mutualistic plant bacterial associations could be predicted via ML algorithms using photosynthetic pigment variations. It is noteworthy to mention the photosynthetic pigment variations flaunted by phytoplankton in response to climatic changes due to anthropogenic activities (Zhang et al., 2019). ML could be successfully used to predict the correlation between climatic fluctuations and pigments in plants residing in different geographical areas and identify targets for genetically modifying pigment content to fit the crop's growing environment. The use of DL methods in the genomics of photosynthetic pigments has yet to be carried out. DL could be employed to determine promoters, enhancers, splicing regulators and their targets, TFs, and RNA-binding proteins, and for structural classification of proteins and secondary structure prediction (Yue & Wang, 2018). The successful implementation of DL in omics techniques such as genomics, proteomics, and metabolomics may aid in unraveling the biosynthetic pathways of many photosynthetic pigments. ML has the potential for use in the prediction of the following genomic elements:

- Transposable elements (TE) are mobile repetitive elements that can determine some genetic variations (Kim, 2017; Orozco-Arias et al., 2019). Detecting them, as well as learning them, is a complex endeavor and ML is a promising tool for these applications. Interestingly, plant TEs are proven to have some unparallel roles in stress and pathogen invasion. SVMs (Support Vector Machines) were successfully used in studying TEs and helitrons and other ML algorithms such as

HMM (Hidden Markov model) and neural networks could also be effectively used to investigate these elements in plants.

- QTLs (Quantitative Trait Loci) are specific genomic regions that are linked to a phenotype. Identification of QTLs is an integral part of genomics studies and RF (Random Forest), SVM and DL are useful algorithms for QTL prediction (Dijk, 2021). Unfortunately, no relevant breakthroughs have been reported regarding pigment studies.

- Similarly, SNP (Single Nucleotide Polymorphism) detection (Aono et al., 2020) could be accomplished through ML. SNPs are conserved mutations occurring at a single base in the genome, which can be sufficient to induce significant differences in phenotype. The successful identification of SNPs through ML in polyploid organisms demonstrates the efficacy of this computational approach in genomic prediction and underlines the potential of ML for genomic prediction of SNPs in plants (which often have high ploidy) (Korani et al., 2019).

- ML could also be employed to predict miRNAs. Precise, prompt identification of miRNAs (small non-coding RNAs that silence gene expression) (Miller, 2020) could aid in genomic strategies to elucidate biosynthetic pathways of photosynthetic pigments.

- Long non-coding RNAs (lncRNAs) are another set of RNAs that have been proven to take part in the development and stress responses of plants. Ensemble ML approaches have previously been used to detect lncRNAs in the plant genome (Simopoulos et al., 2018).

- Gene regulatory networks (GRNs) control metabolic functions and thus their identification could unravel related gene functions in plants. ML algorithms trained on large transcriptome databases could be employed to study GRNs and used to explore pigment biosynthetic pathways (Mochida et al., 2018).

- ML could also be deployed to identify the expression of marker genes in mutational studies on a large scale, generating a novel method to carry out metabolome-based experimentations relating to photosynthetic pigments.

Fig. 5 represents the possible ways to explore ML in photosynthetic pigment research

5. Bottlenecks in ML-driven photosynthetic research

There are still some hurdles to overcome. The adoption of ML in photosynthetic research is still in an underdeveloped stage in many countries. From a practical perspective, the extrapolation of sensing from leaves to the canopy level still faces many challenges such as view angle effects, canopy architecture, and atmospheric effects. While in the ML approach, generating an extensive training data set for complex, heterogeneous data requires time, skills, and resources which limits the application. If training data sets are too small, and not properly constructed spurious 'overfitting' occurs, and the forecasting nature of models is diminished affecting the model (Fu et al., 2019).

Besides, there are several limitations to utilizing hyperspectral reflectance to explore the photosynthetic efficiency of an entire canopy. First, If the spectra range is out of the training dataset, there are chances for imprecise predictions. Second, photosynthesis-unrelated compounds can absorb spectra of specified wavelength generating inaccurate predictions. Third, a model should apply to all species and genotypical variations in it for assessing photosynthesis. Fourth, large variations are observed in ML models built using leaf and canopy spectral data (Fu et al., 2022). However, constructing a universal model considering all the plant varieties and compounds with multivariate spectral properties requires extensive research. This raises the significance of mechanistic models questioning the reliability of ML algorithms. In comparison to ML models, mechanistic models require only small datasets, generate novel hypotheses based on observations, and work on simple mathematical simulations (Barker et al., 2018). Mechanistic models centered on more logical assumptions can predict out of the original data set, while ML models are always limited within the training dataset. These deductive models could be extrapolated to analyze photosynthetic activity on a large scale by utilizing in-vitro experimental observations of the spectral properties of chlorophyll in a single leaf. Unfortunately, mechanistic models have not been much explored in the plant domain. Optimistically, a universal model could be built by integrating both ML and mechanistic models which will be achieved in the near future.

6. Conclusion

Engineering photosynthetic pathways to achieve improved crop yield is subject to intense research. The introduction of hyperspectral techniques has filled the gaps in photosynthesis studies between the plant cell and the global level. The discrepancies in correlating these hyperspectral data with photosynthetic parameters to improve crop yield could be mitigated through various machine learning algorithm-based systems. However, the ML approaches in photosynthesis research are still in the budding stage. Significant investigations should be conducted in ML-based photosynthetic research targeting bolstered plant biomass and productivity. Efficient ML programs to predict the responses of plants to stresses, drastic climate changes, shortage of water and nutrition should be implemented at a global scale. A controlled terrestrial environment maintaining optimum growth conditions for the growth and improved yield of crop plants could be achieved in the near future. We put forward machine learning as a channel for a vibrant agricultural sector with increased productivity through Photosynthesis and photosynthetic pigment research.

Acknowledgements

N.H.D was supported by 'Realising increased photosynthetic efficiency to increase strawberry yields' (BBSRC, BB/S507192/1) awarded to A.J.S. A.J.S is supported by the Growing Kent and Medway Program, UK; Ref 107139. R.V., A. K. C., C.G.P.D, and S.R. thanks VIT management for the support.

Declaration of interests

The authors declare no conflicting interests

CRediT authorship contribution statement

R. V. and S. R. contributed to conceptualizing and writing the manuscript. N.H.D, A. J. S, A. K. C, C. G. P. D contributed to writing and reviewing the manuscript. All authors contributed to the final version of the manuscript.

References

1. A. A. Millar, The function of miRNAs in plants, Plants 9 (2020) 198, https://doi.org/10.3390/plants9020198.
2. A. A. Raines, The Calvin cycle revisited. Photosynth. Res. 75 (2003) 1-10.
3. A. Chlingaryan, S. Sukkarieh, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. Comput. Electron. Agric. 151 (2018) 61–69, https://doi.org/10.1016/j.compag.2018.05.012.
4. A. D. J. Dijk, G. Kootstra, W. Kruijer, D. de Ridder, Machine learning in plant science and plant breeding. iScience 24 (2021) 101890, https://doi.org/10.1016/j.isci.2020.101890.
5. A. Dey, Machine learning algorithms: A review. Int. J. Comput. Sci. Inf. Technol. 7 (2016) 1174–1179.
6. A. H. Aono, E. A. Costa, H. V. S. Rody, J. S. Nagai, R. José, G. Pimenta, M. C. Mancini, F. R. C. dos Santos, L. R. Pinto, M. G. A. landell, A. P. de Souza, R. M. Kuroshu, Machine learning approaches reveal genomic regions associated with sugarcane brown rust resistance. Sci. Rep. 10 (2020) 1–16, https://doi.org /10.1038/s41598-020-77063-5.
7. A. J. Simkin, Genetic engineering for global food security: Photosynthesis and biofortification. Plants 8 (2019) 586, https://doi.org/10.3390/plants8120586.
8. A. J. Simkin, L. Kapoor, C. G. P. Doss, T. A. Hofmann, T. Lawson, S. Ramamoorthy. The role of photosynthesis related pigments in light harvesting, photoprotection and enhancement of photosynthetic yield in planta. Photosynth Res. 152 (2022) 23-42, https://doi.org/10.1007/s11120-021-00892-6.

9. A. J. Simkin, L. McAusland, L. R. Headland, T. Lawson, C. A. Raines. Multigene manipulation of photosynthetic carbon assimilation increases CO_2 fixation and biomass yield in tobacco. J. Exp. Bot. 13 (2015), 4075 – 4090.

10. A. J. Simkin, L. McAusland, T. Lawson, C. A. Raines, Over-expression of the RieskeFeS protein increases electron transport rates and biomass yield. Plant Physiol. 175 (2017a) 134–145, https://doi.org/10.1104/pp.17.00622.

11. A. J. Simkin, P. E. López-Calcagno, C. A. Raines, Feeding the world: Improving photosynthetic efficiency for sustainable crop production. J. Exp. Bot. 70 (2019) 1119–1140, https://doi.org/10.1093/jxb/ery445.

12. A. J. Simkin, P. E. López-Calcagno, P. A. Davey, L. R. Headland, T. Lawson, S. Timm, H. Bauwe, C. A. Raines, Simultaneous stimulation of sedoheptulose 1, 7- bisphosphatase, fructose 1, 6-bisphophate aldolase and the photorespiratory glycine decarboxylase- H protein increases CO2 assimilation, vegetative biomass, and seed yield in Arabidopsis. Plant biotechnol J. 15 (2017b) 805 – 816, https://doi.org/10.1111/pbi.12676.

13. A. Krieger-Liszkay, C. Fufezan, A. Trebst, Singlet oxygen production in photosystem II and related protection mechanism. Photosynth Res 98 (2008) 551-564. https://doi.org/10.1007/s11120-008-9349-3

14. A. O. Conrad, W. Lei, D-Y. Lee, G-L. Wang, L. Rodriguez-Saona, P. Bonello, Machine learning-based presymptomatic detection of rice sheath blight using spectral profiles. Plant Phenomics 2020 (2020) 8954085, https://doi.org/10.34133/2020/8954085

15. A. P. M. Weber, A. Bar-Even Update: Improving the efficiency of photosynthetic carbon reactions. Plant Physiol. 179 (2019) 803–812, https://doi.org/10.1104/pp.18.01521.

16. A. P. De Souza, S. J. Burgess, L. Doran, J. Hansen, L. Manukyan, N. Maryn, D. Gotarkar, L. Leonelli, K. K. Niyogi, S. P.Long. Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. Science. 377 (2022) 851-854, https://doi.org/10.1126/science.adc9831.

17. A. Singh, S. Jones, B. Ganapathysubramanian, S. Sarkar, D. Mueller, K. Sandhu, K. Nagasubramanian, Challenges and opportunities in machine-augmented plant stress phenotyping, Trends Plant Sci. 26 (2021) 53–69, https://doi.org/10.1016/j.tplants.2020.07.010.

18. A. T. Ayumi, Chlorophyll a oxygenase (CAO) is involved in chlorophyll b formation from chlorophyll a. Proc Natl Acad Sci USA 95 (1998) 12719-12723.

19. A. V. Zubler, J. Y. Yoon, Proximal methods for plant stress detection using optical sensors and machine learning. Biosensors 10 (2020) 193, https://doi.org/10.3390/bios10120193.

20. A. Wolanin, G. Camps-Valls, L. Gómez-Chova, G. Mateo-García, C. van der Tol, Y. Zhang, L. Guanter Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sens. Environ. 225 (2019) 441–457, https://doi.org/10.1016/j.rse.2019.03.002.

21. B. Alhnaity, S. Pearson, G. Leontidis, S. Kollias, Using deep learning to predict plant growth and yield in greenhouse environments. Acta Hortic. 1296 (2020) 425–431, https://doi.org/10.48550/arXiv.1907.00624

22. B. J. Walker, A. VanLoocke, C. J. Bernacchi, D. R. Ort, The costs of photorespiration to food production now and in the future. Annu. Rev. Plant Biol. 67 (2016) 107-129, https://doi.org/10.1146/annurev-arplant-043015-111709

23. B. Raychaudhiri, Remote Sensing of Solar-Induced Chlorophyll Fluorescence at Atmospheric Oxygen Absorption Band Around 760 nm and Simulation of That Absorption in Laboratory. nm and simulation of that absorption in laboratory, IEEE Trans Geosci Remote Sens. 50 (2012) 3908–3914, https://doi.org/10.1109/TGRS.2012.2185503.

24. C. A. Raines, A. P. Cavanagh, A. J. Simkin, Improving carbon fixation. In Photosynthesis in Action (1st edition) Academic Press. (2022) 175-192. https://doi.org/10.1016/B978-0-12-823781-6.00009-5.

25. C. B. Field, V. R. Barros. Climate change 2014 impacts, adaptation and vulnerability: Part A: Global and sectoral aspects: Working group II contribution to the fifth assessment report of the

intergovernmental panel on climate change, eds., Cambridge, UK and New York, NY, USA: Cambridge University Press (2014).

26. C. Le. Quéré, M. R. Raupach, J. G. Canadell, G. Marland, L. Bopp, P. Ciais, T. J. Conway, S. C. Doney, R. A. Feely, P. Foster, P. Fredlingstein, K. Gurney, R. A. Houghton, J. I. House, C. Huntingford, P. E. Levy, M. R. Lomas, J. Majkut, N. Metzl, J. P. Ometto, G. P. Peters, I. C. Prentice, J. T. Randerson, S. W. Running, J. L. Sarmiento, U. Schuster, S. Sitch, T. Takahashi, N. Vivoy Trends in the sources and sinks of carbon dioxide. Nat. Geosci. 2 (2009) 831–836

27. C. M. A. Simopoulos, E. A. Weretilnyk, G. B. Golding, Prediction of plant lncRNA by ensemble machine learning classifiers, BMC Genomics 19 (2018) 1–11, https://doi.org/10.1186/s12864-018-4665-2.

28. C. O. Justice, E. Vermote, J. R. G. Townshend, R. Defries, D. P. Roy, D. K. Hall, V. V. Salomonson, J. L. Privette, G. Riggs, A. Strahler, W. Lucht, R. B. Myneni, Y. Knyazikhin, S. W. Running, R. R. Nenmani, Z. Wan, A. R. Huete, W. van Leeuwen, R. E. Wolfe, L. Giglio, J. Muller, P. Lewis, M. J. Barnsley, The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. IEEE Trans Geosci R. 36 (1998) 1228-1249, https://doi.org/10.1109/36.701075.

29. C. R. Yendrek, T. Tomaz, C. M. Montes, Y. Cao, A. M. Morse, P. J. Brown, L. M. Mclntyre, A. D. B. Leakey, E. A. Ainsworth, High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. Plant physiology 173(1) (2017) 614-626, https://doi.org/10.1104/pp.16.01447

30. D. Dunea, V. Moise, Artificial neural networks as support for leaf area modelling in crop canopies. In Proc. 12th WSEAS Int. Conf. Comput. July 2008 (2008) 440–445.

31. D. Heckmann, U. Scluter, A. P. M. Weber, Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra. Mol. Plant 10 (2017) 878–890, https://doi.org/10.1016/j.molp.2017.04.009.

32. E. A. Ainsworth , S. P. Long. What have we learned from 15 years of free-air CO2 enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO2. New Phytol. 165 (2005) 351-71, https://doi.org/10.1111/j.1469-8137.2004.01224.x.

33. E. A. Ainsworth, S. P. Serbin, J. A. Skoneczka, P. A. Townsend, Using leaf optical properties to detect ozone effects on foliar biochemistry. Photosynth Res 119 (2014) 65–76, https://doi.org/10.1007/s11120-013-9837-y.

34. E. Vogel, M. G. Donat, L. V. Alexander, M, Meinshausen, D. K. Ray, D. Karoly, N. Meinhausen, K. Frieler, The effects of climate extremes on global agricultural yields. Environ. Res. Lett. 14 (2019) 054010, https://doi.org/10.1088/1748-9326/ab154b

35. F. J. Knoll, V. Czymmek, L. O. Harders, S. Hussmann, Real-time classification of weeds in organic carrot production using deep learning algorithms. Comput. Electron. Agric. 167 (2019) 105097.

36. FAO. 2009. How to feed the world in 2050. Food and Agricultural Organisation. https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050. [Accessed 26 April 2023]

37. G. An, M. Xing, B. He, C. Liao, X. Huang, J. Shang, H. Kang, Using machine learning for estimating rice chlorophyll content from in situ hyperspectral data. Remote sens. 12 (2020) 3104, https://doi.org/10.3390/rs12183104.

38. G. Polder, P. M Blok, H. A. C. de Villiers, J. M. van der Wolf, J. Kamp, Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. Front. Plant Sci. 10 (2019) 1–13, https://doi.org/10.3389/fpls.2019.00209.

39. G. Tramontana, M. Jung, C. R. Schwalm, K. Ichii, G. Camps-Valls, B. Ráduly, M. Reichstein, M. A. Arain, A. Cescatti, G. Kiely, L. Merbold, P. Serrano-Ortiz, S. Sickert, S. Wolf, D. Papale, Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences 13 (2016) 4291–4313, https://doi.org/10.5194/bg-13-4291-2016.

40. G. Tramontana, M. Migliavacca, M. Jung, M. Reichstein, T. F. Keenan, G. Camps-Valls, J. Ogee, J. Verrelst, D. Papale, Partitioning net carbon dioxide fluxes into photosynthesis and respiration

using neural networks. Glob. Chang. Biol. 26 (2020) 5235–5253, https://doi.org/10.1111/gcb.15203.

41. H. Chida, A. Nakazawa, H. Akazaki, T. Hirano, K. Suruga, M. Ogawa, T. Satoh, K. Kadokura, S. Yamada, W. Hakamata, K. Isobe, T. Ito, R. Ishii, T. Nishio, K. Sonoike, T. Oku, Expression of the algal cytochrome c6 gene in Arabidopsis enhances photosynthesis and growth. Plant Cell Physiol. 48 (2007) 948–957, https://doi.org/10.1093/pcp/pcm064.

42. H. Croft, J.M. Chen, X. Luo, P. Bartlett, B. Chen, R. M. Staebler, Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Glob. Change Biol. 23 (2017) 3513–3524. doi: 10.1111/gcb.13599

43. H. Hashimoto, C. Uragami, R. J. Cogdell Carotenoids and photosynthesis. In: Carotenoids in Nature. Springer, (2016) pp 111-139.

44. H. K. Ledford, K. K. Niyogi Singlet oxygen and photo-oxidative stress management in plants and algae. Plant Cell Environ. 28 (2005) 1037-1045. https://doi.org/10.1111/j.1365-3040.2005.01374.x

45. H. Zhang, S. Huo, K. M. Yeager, Z. He, B. Xi, X. Li, C. Ma, F. Wu, Phytoplankton response to climate changes and anthropogenic activities recorded by sedimentary pigments in a shallow eutrophied lake. Sci. Total Environ. 647 (2020) 1398–1409, https://doi.org/10.1016/j.scitotenv.2018.08.081.

46. I. Domonkos, M. Kis, Z. Gombos, B. Ughy, Carotenoids, versatile components of oxygenic photosynthesis. Progress in lipid research 52 (2013) 539-561. https://doi.org/10.1016/j.plipres.2013.07.001

47. I. E. Naqa, M. J. Murphy, What is machine learning? In Machine learning in radiation oncology: Theory and Applications. (2015) pp. 3–11, Springer

48. IPCC. (2007) Climate Change 2007 - The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC, Cambridge, UK and New York, NY, USA

49. J. B. Féret, G. le Maire, S. Jay, D. Berveiller, R. Bendouls, G. Hmimina, A. Cheraiet, J. C. Oliveira, F. J. Ponzoni, T. Solanki, F. de Boissieu, J. Chave, Y. Nouvellon, A. Porcar-Castell, M.-J. Lefèvre-Fonollosa, Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. Remote Sens. Environ. 231 (2019) 1–82, https://doi.org/10.1016/j.rse.2018.11.002.

50. J. Baron, D. J. Hill, H. Elmiligi, Combining image processing and machine learning to identify invasive plants in high-resolution images. Int. J. Remote Sens. 39 (2018) 5099–5118, https://doi.org/10.1080/01431161.2017.1420940.

51. J. Dong, N. Gruda, X. Li, Y. Tang, P. Zhang, Z. Duan, Sustainable vegetable production under changing climate: The impact of elevated CO2 on yield of vegetables and the interactions with environments-A review. Journal of Cleaner Production 253 (2020) 119920. https://doi.org/10.1016/j.jclepro.2019.119920.

52. J. GU, Z. Zhou, Z. Li, Y. Chen, Z. Wang, H. Zhang Rice (Oryza sativa L.) with reduced chlorophyll content exhibit higher photosynthetic rate and efficiency, improved canopy 1 light distribution, and greater yields than normally pigmented plants. Field Crops Res. 200 (2017) 58-70, https://doi.org/10.1016/j.fcr.2016.10.008.

53. J. Huang, J. Zheng, H. Yuan, K. McGinnis, Distinct tissue-specific transcriptional regulation revealed by gene regulatory networks in maize. BMC Plant Biol. 18 (2018) 1–14, https://doi.org/10.1186/s12870-018-1329-y.

54. J. -J. Zhou, Y. -H. Zhang, Z. -M. Han, X. -Y. Liu, Y. -F. Joan, C. -G. Hu, Y. -Y. Dian, Evaluating the Performance of Hyperspectral Leaf Reflectance to Detect Water Stress and Estimation of Photosynthetic Capacities. Remote Sens. 13 (2021) 2160, https://doi.org/10.3390/rs13112160.

55. J. J. Zhou, Y. -H. Zhang, Z. -M. Han, Y. -Y Liu, Y. -F. Jian, C. -G. Hu, Y.-Y. Dian, Hyperspectral sensing of photosynthesis, stomatal conductance, and transpiration for citrus tree under drought condition. Bio Rxiv, (2021), https://doi.org/10.1101/2021.02.26.433135.

56. J. Kromdijk, K. Glowacka, L. Leonelli , S. T. Gabilly, M. Iwai, K. K. Niyogi, S. P. Long, Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science 354 (2016) 857-861. https://doi.org/10.1126/science.aai8878

57. J. M. Gonalez- Camacho, L. Ornella, P. Perez-Rodriguez, D. Gianola, S. Dreisigacker, J. Crossa, Applications of machine learning methods to genomic selection in breeding wheat for rust resistance. Plant Genome 11 (2018) 170104.

58. J. Wu, A. Rogers, L. P. Albert, K. Ely, N. Prohaska, B. T. Wolfe, R. C. Oliveira Jr, S. R. Saleska, S. P. Serbin, Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests, New Phytol. 224 (2019) 663-674, https://doi.org/10.1111/nph.16029.

59. K. Zheng, Y. Bu, Y. Bao, X. Zhu, J. Wang, Y. Wang, A machine learning model for photorespiration response to multi-factors. Horticulturae 7 (2021) 207, https://doi.org/10.3390/horticulturae7080207

60. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis, Machine learning in agriculture: A review. Sensors 18 (2018) 1–29.

61. K. K. Niyogi, X. P. Li, V. Rosenberg, H-S Jung, Is PsbS the site of non-photochemical quenching in photosynthesis? J. Exp. Bot. 56 (2004) 375-382. https://doi.org/10.1093/jxb/eri056.

62. K. Meacham-Hensold, C. M. Montes, J. Wu, K. Guan, P. Fu, E. A. Ainsworth, T. Pederson, C. E. Moore, K. L. Brown, C. Raines, C. J. Bernacchi, High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sensing of Environment 231 (2019)111176, https://doi.org/10.1016/j.rse.2019.04.029.

63. K. Mochida, S. Koda, K. Inoue, R. Nishii, Statistical and machine learning approaches to predict gene regulatory networks from transcriptome datasets. Front. Plant Sci. 871 (2018) 1–7, https://doi.org/10.3389/fpls.2018.01770.

64. K. Przybył, J. Wawrzyniak, K. Koszela, F. Adamski, M. Gawrysiak-Witulska, Application of deep and machine learning using image analysis to detect fungal contamination of rapeseed. Sensors (Switzerland) 20 (2020) 1–11, https://doi.org/10.3390/s20247305.
65. K. Turgut, H. Dutagaci, G. Galopin, D. Rousseau, Segmentation of structural parts of rosebush plants with 3D point-based deep learning methods. Plant Methods, BioMed Central 18 (2022) 20, https://doi.org/10.1186/s13007-022-00857-3
66. K. Yamamoto, W. Guo, Y. Yoshioka, S. Ninomiya, On plant detection of intact tomato fruits using image analysis and machine learning methods. Sensors (Switzerland) 14 (2014) 12191–12206, https://doi.org/10.3390/s140712191.
67. Kim, N. S. (2017) The genomes and transposable elements in plants: are they friends or foes? Genes and Genomics 39, 359–370, https://doi.org/10.1007/s13258-017-0522-y
68. L. Arnold, S. Rebecchi, S. Chevallier, H. Paugam-Moisy, An introduction to deep learning. In Eur. Symp. Artif. Neural Networks (2011) 477–488.
69. L. Cotrozzi, R. Peron, M. R. Tuinstra, M. V. Mickelbart, J. J. Couture, Spectral Phenotyping of Physiological and Anatomical Leaf Traits Related with Maize Water Status. Plant Physiol. 184 (2020) 1363-1377, https://doi.org/10.1104/pp.20.00577
70. L. E. O. Aparecido, G. S. Rolim, J. R. S. C. Moraes, C. T. S. Costa, P. S. de Souza, Machine learning algorithms for forecasting the incidence of Coffea arabica pests and diseases. Int. J. Biometeorol. 64 (2020) 671–688, https://doi.org/10.1007/s00484-019-01856-1.
71. L. Han, G. Yang, H. Dai, B. Xu, H. Yang, H. Feng, Z. Li, X. Yang, Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. Plant Methods 15 (2019) 1–19, https://doi.org/10.1186/s13007-019-0394-z.
72. L. Li, H. Yi, Photosynthetic responses of Arabidopsis to SO2 were related to photosynthetic pigments, photosynthesis gene expression and redox regulation. Ecotoxicol. Environ. Saf. 203 (2020) 111019, https://doi.org/10.1016/j.ecoenv.2020.111019.

73. L. M. Mortensen, Effects of elevated CO2 concentrations on growth and yield of eight vegetable species in a cool climate. Sci. Hortic. 58 (1994) 177-185. https://doi.org/10.1016/0304-4238(94)90149-X (1994).

74. L. Mkonyi, D. Rubanga, M. Richard, N. Zekeya, S. Sawahiko, B. Maiseli, D. Machuve, Early identification of Tuta absoluta in tomato plants using deep learning. Sci. African 10 (2020) e00590, https://doi.org/10.1016/j.sciaf.2020.e00590.

75. L. P. Osco, A. P. M. Ramos, É. A. S. Moriya, L. G. Bavaresco, B. C. de Lima, N. Estrabis, D. R. Pereira, J. E. Creste, J. M. Junior, W. N. Gonçalves, N. N. Imai, J. Li, V. Liesenberg, F. F. de Araújo, Modeling Hyperspectral Response of Water-Stress Induced Lettuce Plants Using Artificial Neural Networks. Remote. Sens. 11 (2019) 2797, https://doi.org/10.3390/rs11232797.

76. L. P. Osco, A. P. M. Ramos, M. M. F. Pinheiro, E. A. S. Moriya, N. N. Imai, N. Estrabis, F. Ianczyk, F. F. de Araújo, V. Liesenberg, L. A. C. Jorge, J. Li, L. Ma, W. N. Gonçalves, J. M. Junior, J. E. Creste, A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements. Remote Sens. 12 (2020) 906, https://doi.org/10.3390/rs12060906.

77. L. T. Evans, R. A, Fischer Yield potential: Its definition, measurement, and significance. Crop Sci. 39 (1999) 1544–1551, https://doi.org/10.2135/cropsci1999.3961544x.

78. M. Amin. Predicting the oxidation states of Mn ions in the oxygen-evolving complex of photosystem II using supervised and unsupervised machine learning. Photosynth. Res. 156 (2022) 89-100. https://doi.org/10.1007/s11120-022-00941-8.

79. M. Ermakova, P. E. López-Calcagno, C. A. Raines, R. T. Furbank, S. von Caemmerer, Overexpression of the Rieske FeS protein of the Cytochrome b6f complex increases C4 photosynthesis in Setaria viridis. Commun. Biol. 2 (2019) 314, https://doi.org/10.1038/s42003-019-0561-9.

80. M. Forkel, M. Drüke, M. Thurner, W. Dorigo, S. Schaphoff, K. Thonicke, W. von Bloh, N. Carvalhais, Constraining modelled global vegetation dynamics and carbon turnover using multiple satellite observations, Sci. Rep. 9 (2019) 1–12, https://doi.org/10.1038/s41598-019-55187-7.

81. M. H. Siebers, N. Gomez-Casanovas, P. Fu, K. Meacham-Hensold, C. E. Moore, C. J. Bernacchi, Emerging approaches to measure photosynthesis from the leaf to the ecosystem, Emerg Top Life Sci. 5 (2021) 261-274, https://doi.org/10.1042/ETLS20200292.

82. M. Hesami, M. Alizadeh, A. M. P. Jones, D. Torkamaneh, Machine learning: Its challenges and opportunities in plant system biology. Appl. Microbiol. Biotechnol. 106 (2022) 3507-3530.

83. M. I. Jordan, T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. Science. 349 (2015) 255-260.

84. M. Schlund, V. Eyring, G. Camps-Valls, P. Friedlingstein, P. Gentine, M. Reichstein, Constraining uncertainty in projected gross primary production with machine learning, J. Geophys. Res. Biogeosciences 125 (2020) 1–22, https://doi.org/10.1029/2019JG005619.

85. M.L. Buchaillot, D. Soba, T. Shu, J. Liu, I. Aranjuelo, J. L. Araus, G. B. Runion, S. A. Prior, S. C. Kefauver, A. Sanz-Saez, Estimating peanut and soybean photosynthetic traits using leaf spectral reflectance and advance regression models. Planta 25 (2022) 93, https://doi.org/10.1007/s00425-022-03867-6

86. M.M. Ozguven, K. Adem, Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms. Phys. A Stat. Mech. its Appl. 535 (2019) 122537.

87. N. Friedland, S. Negi, T. Vinogradova-Shah, G. Wu, L. Ma, S. Flynn, T. Kumssa, C-H Lee, R. Sayre, Fine-tuning the photosynthetic light harvesting apparatus for improved photosynthetic efficiency and biomass yield. Scientific Reports 9 (2019) 1-12. https://doi.org/10.1038/s41598-019-49545-8.

88. N. H. Doddrell, T. Lawson, C. A. Raines, C. Wagstaff, A. J. Simkin, Feeding the world: impacts of elevated [CO_2] on nutrient content of greenhouse grown fruit crops and options for future yield gains. Horticultural Research 10 (2023) 26. https://doi.org/10.1093/hr/uhad026

89. O. Dermody, S. P. Long, K. Mcconnaughay, E. H. DeLucia, How do elevated CO_2 and O3 affect the interception and utilization of radiation by a soybean canopy? Glob. Chang. Biol. 14 (2008) 556–564, https://doi.org/10.1111/j.1365-2486.2007.01502.x

90. P. B. Morgan, G. A. Bollero, R. L. Nelson, F. G. Dohleman, S. P. Long, Smaller than predicted increase in above ground net primary production and yield of field-grown soybean under fully open-air [CO_2] elevation. Glob. Chang. Biol. 11 (2005) 1856–1865, https://doi.org/10.1111/j.1365-2486.2005.001017.x.

91. P. E. López-Calcagno, K. L. Brown, A. J. Simkin, S. J. Fisk, S. Vialet-Chabrand, T. Lawson, C. A. Raines, Stimulating photosynthetic processes increases productivity and water-use efficiency in the field. Nature Plants. 6 (2020) 1054–1063.

92. P. E. López-Calcagno, S. Fisk, K. L. Brown, S. E. Bull, P. F. South, C. A. Raines, Overexpressing the H- protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. Plant Biotechnol J. 17 (2018) 141– 151, https://doi.org/10.1111/pbi.12953.

93. P. F. South, A. P. Cavanagh, P. E. López-Calcagno, C. A. Raines, D. R. Ort Optimizing photorespiration for improved crop productivity. J. Integr. Plant Biol. 60 (2018) 1217–1230, https://doi.org/10.1111/jipb.12709.

94. P. Fu, C. M. Montes, M. H. Siebers, N. Gomez-Casanovas, J. M. McGrath, E. A. Ainsworth, C. J. Bernacchi, Advances in field-based high-throughput photosynthetic phenotyping, J. Exp. Bot. 73 (2022) 3157-3172, https://doi.org/10.1093/jxb/erac077.

95. P. Fu, K. Meacham-Hensold, K. Guan, C. J. Bernacchi, Hyperspectral Leaf Reflectance as Proxy for Photosynthetic Capacities: An Ensemble Approach Based on Multiple Machine Learning Algorithms, Front. Plant Sci 10 (2019) 1–13, https://doi.org/10.3389/fpls.2019.00730.

96. P. Fu, K. Meacham-Hensold, K. Guan, J. Wu, C. Bernacchi, Estimating photosynthetic traits from reflectance spectra: A synthesis of spectral indices, numerical inversion, and partial least square regression. Plant Cell Environ. 43 (2020) 1241–1258, https://doi.org/10.1111/pce.13718. Epub 2020 Feb 27.

97. P. Fu, K. Meacham-Hensold, M. H. Siebers, C. J. Bernacchi, The inverse relationship between solar-induced fluorescence yield and photosynthetic capacity: Benefits for field phenotyping. J. Exp. Bot. 72 (2021) 1295–1306, https://doi.org/10.1093/jxb/eraa537

98. P. Garcia-Perez, E. Lozano-Milo, M. Landin, P. P. Gallego, Machine learning unmasked nutritional imbalances on the medicinal plant *Bryophyllum* sp. cultured *in vitro.* Front. Plant Sci. 11 (2020) 1–14, https://doi.org/10.3389/fpls.2020.576177

99. P. Zhang, G. Zhilng, S. Ullah, G. Melagraki, A. Afantitis, I. Lynch. Nanotechnology and artificial intelligence to enable sustainable and precision agriculture, Nature Plants. 7 (2021) 864 – 876.

100. Q. Zhou, A. fellows, G. N. Flerchinger, A. N. Flores, Examining interactions between and among predictors of net ecosystem exchange: A machine learning approach in a semi-arid landscape. Sci. Rep. 9 (2019) 1–11, https://doi.org/10.1038/s41598-019-38639-y.

101. R. E. Baker, J. -M. Peña, J. Jayamohan, A. Jérusalem, Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biol. Lett. 14 (2018) 20170660, https://doi.org/10.1098/rsbl.2017.0660.

102. R. Sonobe, Y. Hirono, A. Oi, Non-destructive detection of tea leaf chlorophyll content using hyperspectral reflectance and machine learning algorithms. Plants 9 (2020) 368, https://doi.org/10.3390/plants9030368

103. R. T. Furbank, J. A. Jimenez-Berni, B. George-Jaeggli, A. B. Potgieter, D. M. Deery, Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. New Phytol. 223 (2019) 1714–1727, https://doi.org/10.1111/nph.15817

104. R. T. Furbank, R. Sharwood, G. M. Estavillo, V. Silva-Perez, A. G. Condon, Photons to food: Genetic improvement of cereal crop photosynthesis, J. Exp. Biol. 71 (2020) 2226–2238, https://doi.org/10.1093/jxb/eraa077.

105. R. T. Furbank, V. Silva-Perez, J. R. Evans, A. G Wheat physiology predictor: predicting physiological traits in wheat from hyperspectral reflectance measurements using deep learning, Plant Methods. 17 (2021) 108, https://doi.org/10.1186/s13007-021-00806-6.

106. S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao. M. P. Reyes, M. Shyu, S. Chen, S. S. Iyengar, A survey on deep learning: Algorithms, techniques, and application. In ACM Computing Surveys 51 (2018) 92:1–92:33, 0 https://doi.org/10.1145/3234150.

107. S. Asseng, F. Ewert, P. Martre, R. P. Rötter, D. B. Lobell, D. Cammarano, B. A. Kimball, M. J. Ottman, G. W. Wall, J. W. White, M. P. Reynolds, P. D. Alderman, P. V. V. Prasad, P. K. Aggarwal, J. Anothai, B. Basso, C. Biernath, A. J. Challinor, G. De Sanctis, J. Doltra, E. Fereres, M. Garcia-Vila, S. Gayler, G. Hoogenboom, L. A. Hunt, R. C. Izaurralde, M. Jabloun, C. D. Jones, K. C. Kersebaum, A-K. Koehler, C. Müller, S. Naresh Kumar, C. Nendel, G. O'Leary, J. E. Olesen, T. Palosuo, E. Priesack, E. Eyshi Rezaei, A. C. Ruane, M. A. Semenov, I. Shcherbak, C. Stöckle, P. Stratonovitch, T. Streck, I. Supit, F. Tao, P. J. Thorburn, K. Waha, E. Wang, D. Wallach, J. Wolf, Z. Zhao, Y. Zhu, Rising temperatures reduce global wheat production, Nat. Clim. Chang. 5 (2015) 143–147, https://doi.org/10.1038/nclimate2470.

108. S. Bhadra, V. Sagan, M. Maimaitijang, M. Maimaitiyiming, M. Newcomb, N. Shakoor, T. C. Mockler, Quantifying leaf chlorophyll concentration of sorghum from hyperspectral data using derivative calculus and machine learning. Remote Sens. 12 (2020) 2082, https://doi.org/10.3390/rs12132082.

109. S. Brown, Machine learning, explained. MIT Management Sloan School. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained. [Accessed 30 June 2023].

110. S. H. Alemohammad, B. Fang, A. G. Konings, F. Aries, J. K. Green, J. Kolassa, D. Miralles, C. Prigent, P. Gentine, Energy, and Carbon with Artificial Neural Networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence, Biogeosciences. 14 (2017) 4101-4124, https://doi.org/10.5194/bg-14-4101-2017.

111. S. K. Yadav, K. Khatri, M. S. Rathore, B. Jha, Introgression of UfCyt c6, a thylakoid lumen protein from a green seaweed Ulva fasciata Delile enhanced photosynthesis and growth in tobacco. Mol. Biol. Rep. 45 (2018) 1745–1758, https://doi.org/10.1007/s11033-018-4318-1.

112. S. Lefebvre, T. Lawson, M. Fryer, O. V. Zakheleniuk, J. C. Llyod, C. A. Raines, Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. Plant Physiol. 138 (2005) 451–60, https://doi.org/10.1104/pp.104.055046

113. S. M. Driever, A. J. Simkin, S. Alotaibi, S. J. Fisk, P. J. Madgwick, C. A. Sparks, H. D. Jones, T. Lawson, M. A. J. Parry, C. A. Raines, Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. Philos. Trans. R. Soc. B Biol. Sci. 372 (2017) 1–10, https://doi.org/10.1098/rstb.2016.0384.

114. S. Orozco- Arias, G. Isaza, R. Guyot, Retrotransposons in plant genomes: Structure, identification, and classification through bioinformatics and machine learning, Int. J. Mol. Sci. 20 (2019) 1–31, https://doi.org/10.3390/ijms20153837.

115. S. P. Long, X. G. Zhu, S. L. Naidu, D. R. Ort, Can improvement in photosynthesis increase crop yields? Plant, Cell Env 13 (2020) 315–330.

116. S. S. Khurschev, T. Y. Plyusnina, T. K. Antal, S. I. Pogosyan, G. Y. Riznichenko, A. B. Rubin, Machine learning methods for assessing photosynthetic activity: environmental monitoring applications. Biophys Rev 14 (2022) 821-842, https://doi.org/10.1007/s12551-022-00982-2.

117. S. Uygun, C. B. Azodi, S-H. Shiu, Cis-regulatory code for predicting plant cell-type transcriptional response to high salinity. Plant Physiol. 181 (2019) 1739–1751, https://doi.org/10.1104/pp.19.00653.

118. S. Wang, S. Guan, Z. Wang, E. A. Ainsworth, T. Zheng, P. A. Townsend, K. Li, C. Moller, G. Wu, C. Jiang, Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy, J. of Exp. Bot 72 (2020) 341-354, https://doi.org/10.1093/jxb/eraa432.

119. T. Liu, Q. Yuan, Y. Wang, Prediction model of photosynthetic rate based on SOPSO-LSSVM for regulation of greenhouse light environment. Eng. Lett. 29 (2021) 297–301.

120. T. Sexton, S. Sankaran, A. B. Cousins, Predicting photosynthetic capacity in tobacco using shortwave infrared spectral reflectance. J. Exp. Bot. 72 (2021) 4373–4383, https://doi.org/10.1093/jxb/erab118.

121. T. U. Kampe, B. R. Johnson, M. Kuester, M. Keller, NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure, J Appl Remote Sens 4(1) (2010) 043510, https://doi.org/10.1117/1.3361375.

122. T. Yue, H. Wang, Deep learning for genomics: A concise overview. In Handbook of Deep Learning Applications, pp. 1–37, Springer

123. V. Mazzia, L. Comba, A. Khaliq, M. Chiaberge, P. Gay, UAV and machine learning based refinement of a satellite-driven vegetation index for precision agriculture. Sensors (Switzerland) 20 (2020) 2530, https://doi.org/10.3390/s20092530

124. V. Silva-Perez, G. Molero, S. P. Serbin, A. G. Cindon, M. P. Reynolds, R. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. J. Exp. Bot. 69 (2018) 483–496, https://doi.org/10.1093/jxb/erx421.

125. W. Korani, J. P. Clevenger, Y. Chu, P. Ozias-Akins, Machine learning as an effective method for identifying true single nucleotide polymorphisms in polyploid plants. Plant Genome 12 (2019) 1–10, 10.3835/plantgenome2018.05.0023

126. W. Wang, W. Kong, T. Shen, Z. Man, W. Zhu, Y. He, F. Liu, Y. Liu, Application of Laser-Induced Breakdown Spectroscopy in Detection of Cadmium Content in Rice Stems. Front. Plant Sci. 11 (2020) 1–11, https://doi.org/10.3389/fpls.2020.599616.

127. W. Wang, Y. Cheng, Y. Ren, Z. Zhang, H. Geng, Prediction of Chlorophyll Content in Multi-Temporal Winter Wheat Based on Multispectral and Machine Learning, Front. Plant Sci. 13 (2022) 896408, https://doi.org/10.3389/fpls.2022.896408.

128. X. E. Pantazi, D. Moshou, T. Alexandridis, R. L. Whetton, A. M. Mouazen, Wheat yield prediction using machine learning and advanced sensing techniques. Comput. Electron. Agric. 121 (2016) 57–65, https://doi.org/10.1016/j.compag.2015.11.018.

129. X. -G. Zhu, S. P. Long, D. R. Ort, Improving photosynthetic efficiency for greater yield. Annu. Rev. Plant Biol. 61 (2010) 235–261, https://doi.org/10.1146/annurev-arplant-042809-112206.

130. X. Xie, A. Li, J. Tan, G. Lei, H. Jin, Z. Zhengjian, Uncertainty analysis of multiple global GPP datasets in characterizing the lagged effect of drought on photosynthesis, Ecol. Indic. 113 (2020) 106224, https://doi.org/10.1016/j.ecolind.2020.106224.

131. X. Xie, J. Shen, Waterlogging Resistance Evaluation Index and Photosynthesis Characteristics Selection: Using Machine Learning Methods to Judge Poplar's Waterlogging Resistance, Mathematics 9 (2021) 1542, https://doi.org/10.3390/math9131542.

132. X. Zhou, Y. Kono, A. Win, T. Matsui, T. S. T. Tanaka, Predicting within-field variability in grain yield and protein content of winter wheat using UAV-based multispectral imagery and machine learning approaches. Plant Prod. Sci. 24 (2021) 137–151, https://doi.org/10.1080/1343943X.2020.1819165

133. Y. Ge, A. Atefi, H. Zhang, C. Miao, R. K. Ramamurthy, B. Sigmon, J. Yang, J. C. Schnable, High-throughput analysis of leaf physiological and chemical traits with VIS–NIR– SWIR spectroscopy: a case study with a maize diversity panel, Plant Methods 15 (2019) 66, https://doi.org/10.1186/s13007-019-0450-8.

134. Y. Li, N.He , J. Hou , L. Xu , C. Liu, J. Zhang, Q. Wang, X. Zhang, X. Wu, Factors Influencing Leaf Chlorophyll Content in Natural Forests at the Biome Scale. Front. Ecol. Evol. 6 (2018) 64. https://doi.org /10.3389/fevo.2018.00064.

135. Y. S. Chung, U. Lee, S. Heo, R. R. Silva, C-I Na, Y. Kim, Image-Based Machine Learning Characterizes Root Nodule in Soybean Exposed to Silicon. Front. Plant Sci. 11(2020) 1–13, https://doi.org/10.3389/fpls.2020.520161

136. Y. Zhang, N. C. Parazoo, A. P. Williams, S. Zhou, P. Gentine, Large and projected strengthening moisture limitation on end-of-season photosynthesis, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 9216–9222, https://doi.org/10.1073/pnas.1914436117

Figure legends

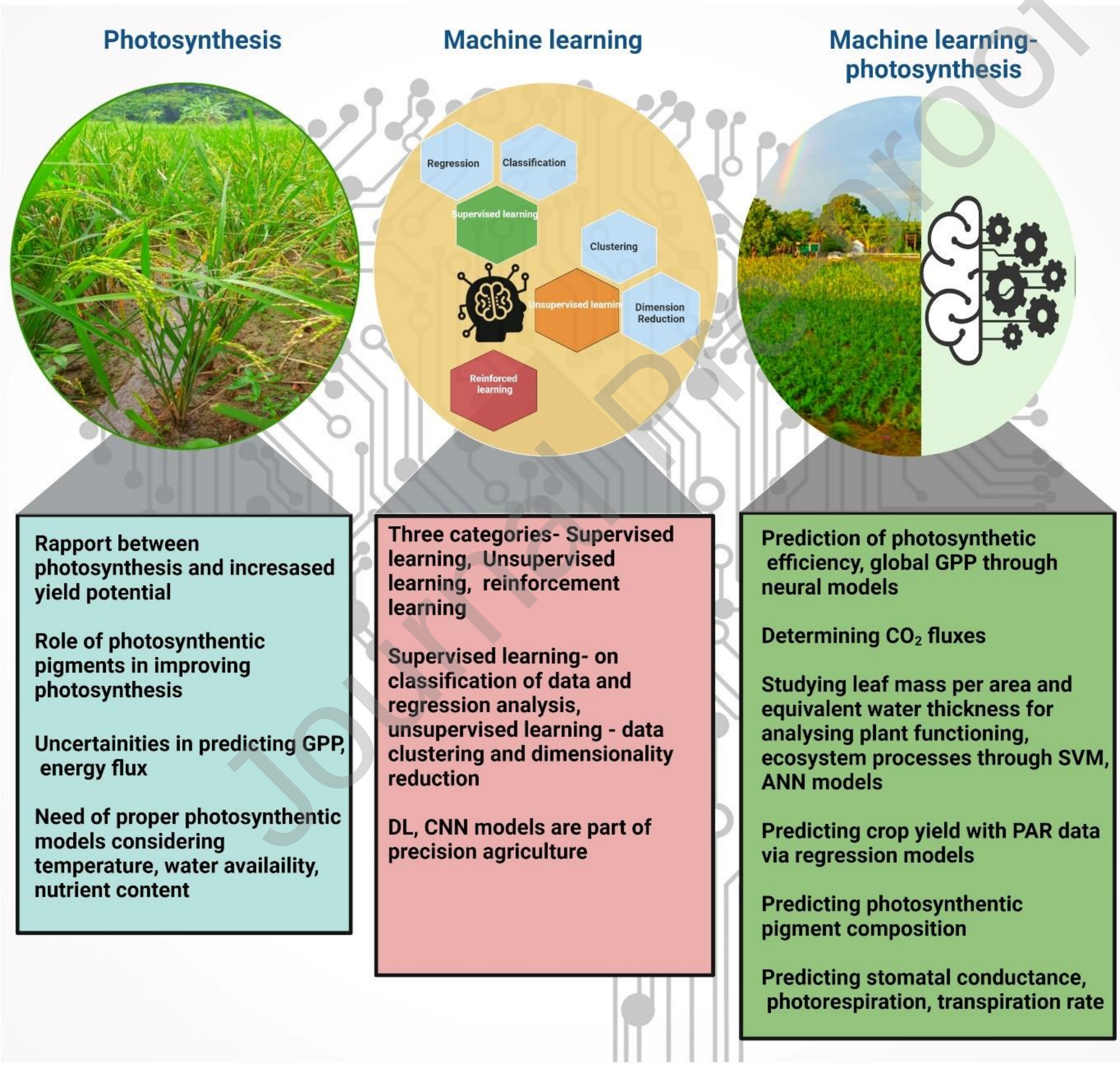

Fig. 1. Overview of ML in photosynthetic research representing the importance of ML approaches and how it could improve photosynthetic research and thus crop yield.

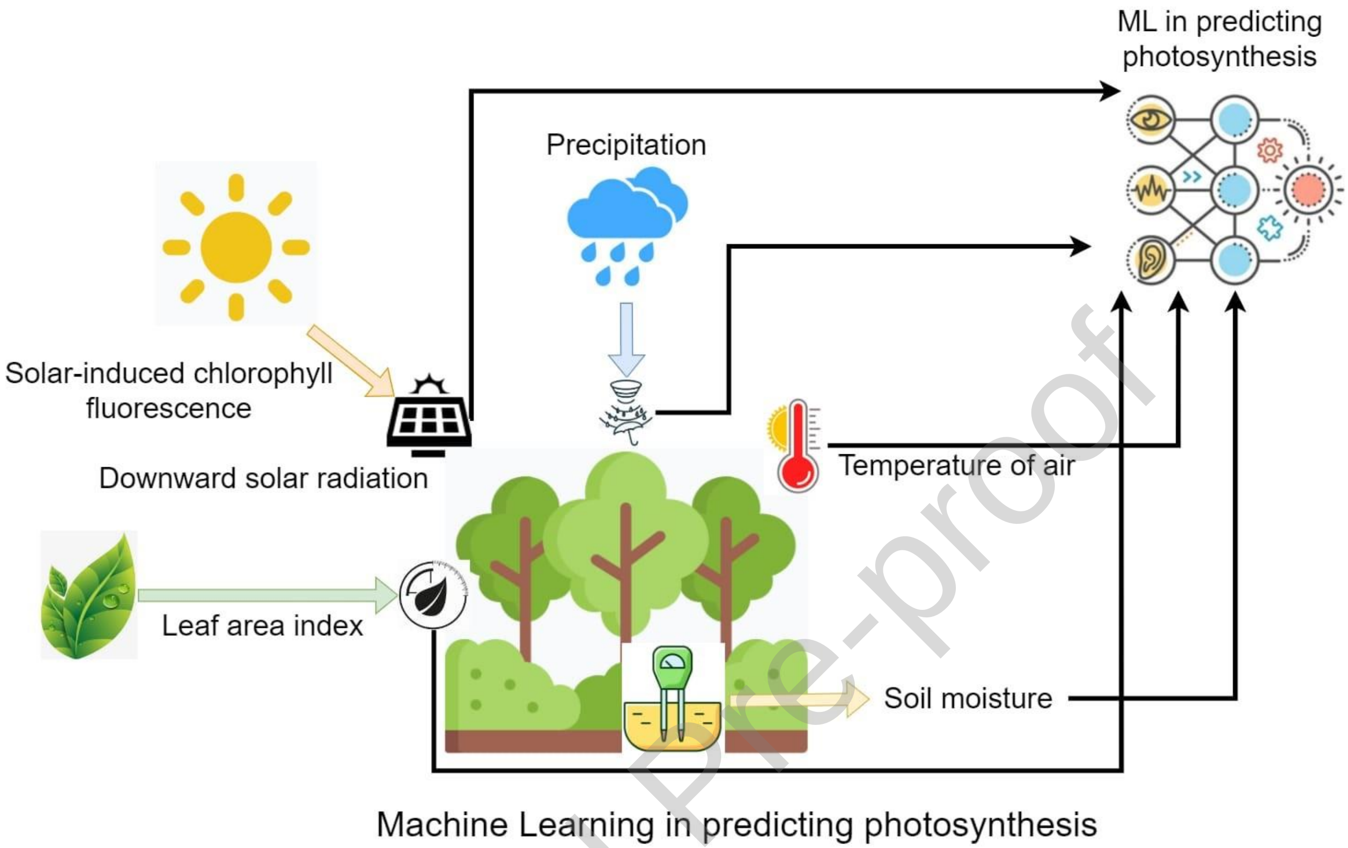

Fig.2. Machine learning in predicting photosynthesis. The data of chlorophyll reflectance from leaves, leaf area index, precipitation rate, temperature variations, soil moisture content are the major parameters used by ML models to predict photosynthetic rate.

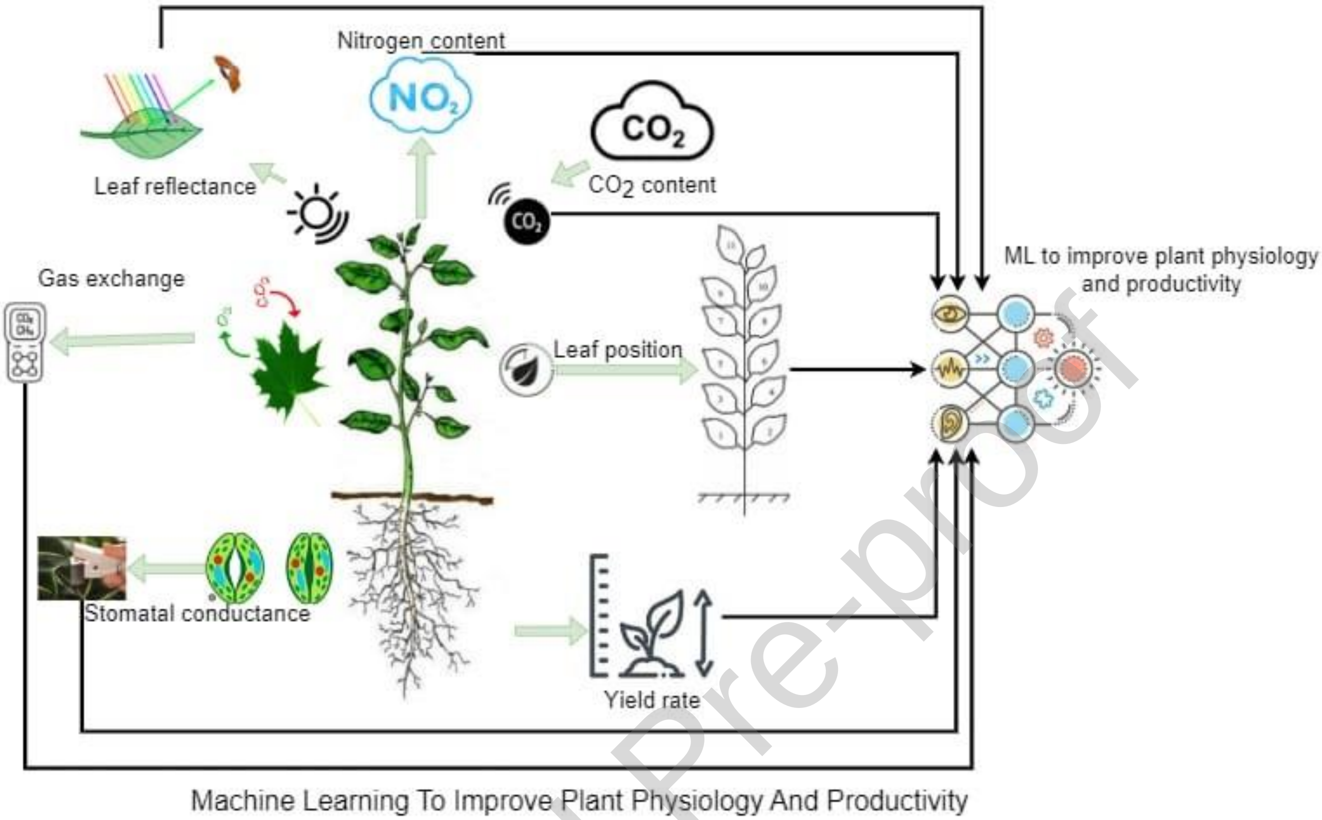

Fig. 3. Machine learning in studying plant physiology and productivity. The CO_2 content, leaf reflectance data, gas exchange rate, stomatal conductance rate, leaf position are some of the parameters generally utilized to assess the plant physiology and productivity using ML algorithms like regression models

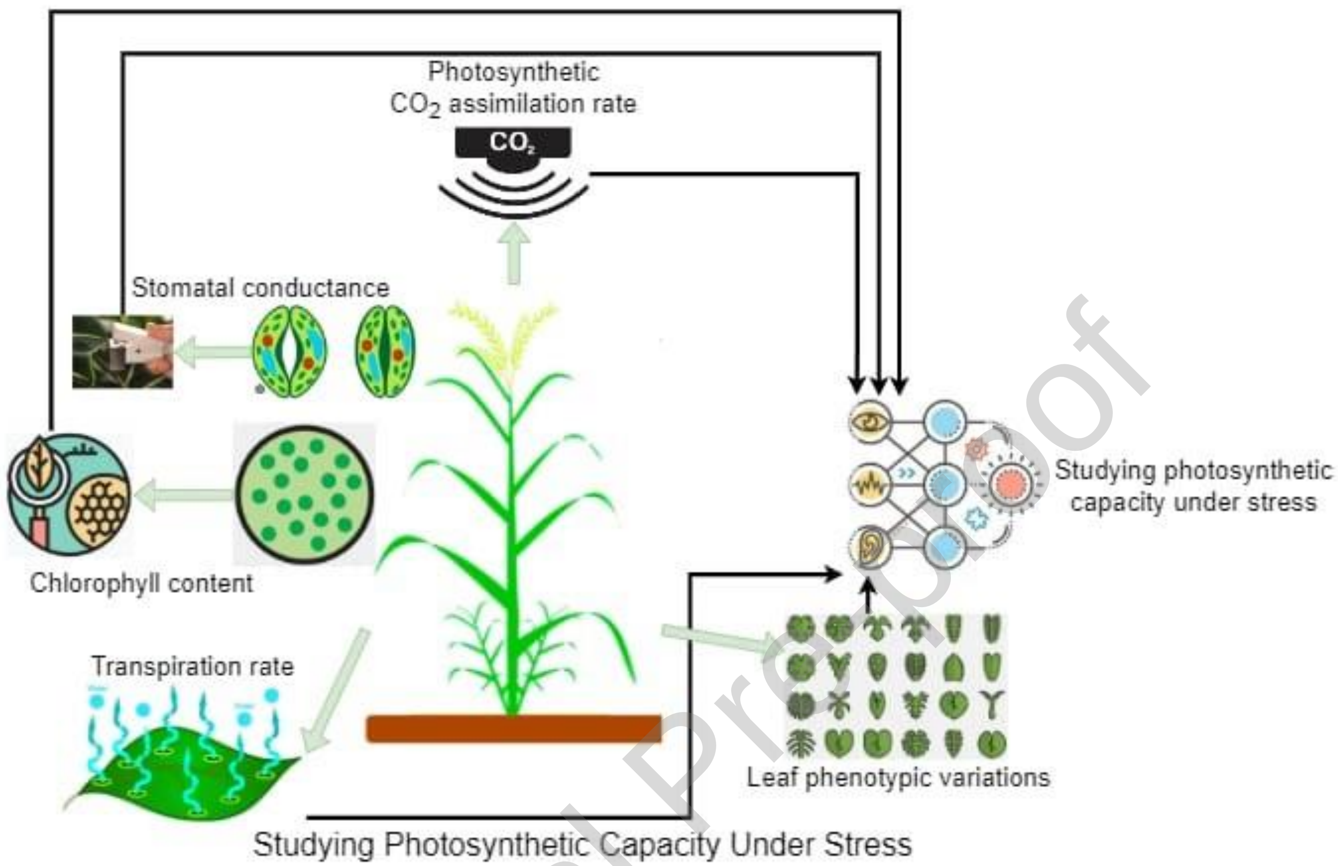

Fig. 4. Machine learning in studying photosynthetic capacity under stress. The data derived from photosynthetic CO_2 assimilation rate, stomatal conductance, chlorophyll content, transpiration rate, leaf phenotypic variations are generally analysed using various models to predict the photosynthetic capacity and productivity during stress conditions like drought.

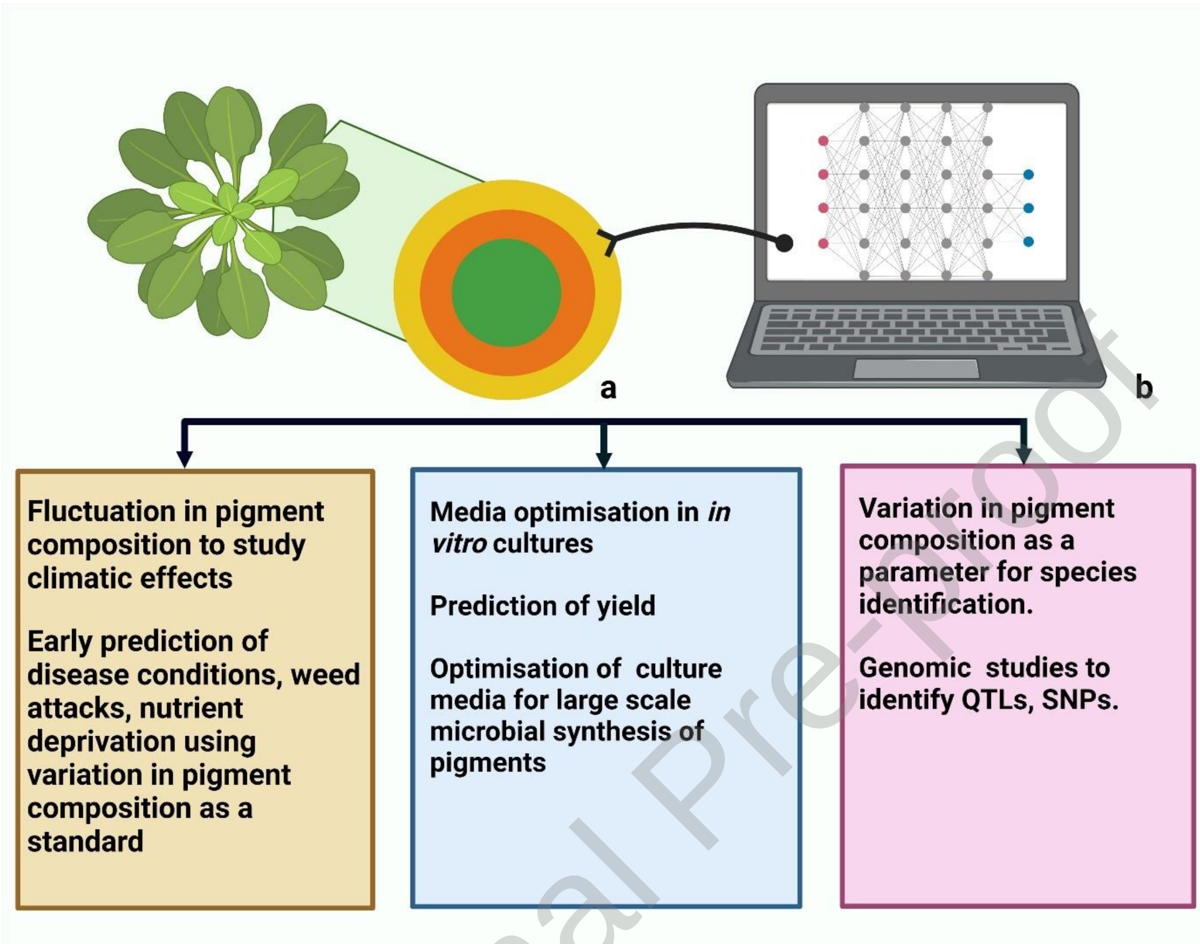

Fig.5. Possible ways to apply machine learning in pigment research a. Pigments- Chlorophyll and accessory pigments, b- ML algorithms. The spectral data derived from plants could be utilized for mentioned applications.

Box 1| Types of machine learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI), which is aimed at building models that exhibit human intelligence behaviour (Brown et al., 2021; Dijk, 2021). The source for building such intelligent models is the data, using which ML models are trained. ML algorithms contain set of instructions that enable the machine to learn from the given data and perform certain tasks that are complex in nature. With such ability they are able to detect patterns, structures in the data, perform tasks such as classification, clustering, prediction, etc. ML algorithms contains five key ingredients including raw data used for training, encoding the features of data, learning from the features, evaluation of learning and optimization for increasing efficiency and accuracy. Once the learning is complete we obtain a machine learning model, using which machine performs the required set of tasks. The ML model represents the learning out of ML algorithm on the training data and hence varies based on the type of ML algorithm. A neural network ML model contains the NN connection weights, bias values etc learned by the NN algorithm on the given data. Similarly, a linear regression model contains vector of coefficients and constants that is best fit for the given training data (Brown et al., 2021).

The data is divided into training data, using which the ML model would be built and testing data, using which the ML model would be evaluated for its accuracy and efficiency. Based on the data available, the type of task to be completed and the style of learning, ML algorithms are classified into following categories (Jordan et al., 2015):

Supervised Learning: The data used for training is labelled and tagged. Learning algorithm optimize its learning by comparing the learned component against the intended output. The best example is classification of plant stress types from chlorophyll fluorescence data (Hesami et al., 2022). Support Vector Machines, Decision Trees, Regression, etc are the well-known supervised learning algorithms.

Unsupervised Learning: The algorithm explores the unlabelled data without any intended output. This exploration leads to derive inferences from the data and identify the latent structures in the data. The best example is clustering of data into multiple groups based on their similarity. K-means clustering, Hierarchical clustering, Gaussian mixture models, etc are the best examples for unsupervised learning algorithms (Jordan & Mitchell, 2015).

Semi Supervised Learning: Algorithms that use a combination of small amount of labelled and large amount of unlabelled data for training purpose. The algorithms build the models that can predict the labels for the instances in the test data and also predict the labels for the unlabelled instances in the training data.

Reinforcement Learning: The algorithm trains the machine to learn from the experiences through interactions with the environment. These interactions provide either rewards or penalties.

Deep Learning: A sub-field of ML algorithms that set the basic parameters describing the data and train the machine to learn via multiple processing layers. During learning, algorithm automatically recognizes the patterns in the data. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) are the best examples of Deep learning (Arnold et al., 2011; Pouyanfar et al., 2018; Brown et al., 2021).

Based on the kind of tasks aimed to solve, ML algorithms are classified into following categories (Jordan &Mitchell, 2015):

Descriptive Learning: ML algorithm aim to explain what has happened from the training data. Unsupervised ML models are deployed for this task.

Predictive Learning: ML algorithms build the ML model that can make predictions about future from the current training data. The best example task is the predicting the class of a new data item, forecast the values such as age, salary, etc. Generally supervised ML models are deployed.

Prescriptive Learning: ML algorithms use the training data to build the model that can make suggestions about the actions that can be taken for the future data. These algorithms work on top of the predictive models and provide recommendations.

Table 1. Application of ML algorithms in different domains of plant science

Plant studied	Algorithm used	Application	Remarks	Reference
Tomato and Ficus	SVR, M5-prime regression trees, RF, K-Nearest Neighbours, LSTM	Yield prediction Stem growth rate	• Framed a DL approach with LSTM • Based on CO2, humidity, radiation, temperature inside and outside green house	(Alhanity et al., 2020)

Maize, wheat, rice, soyabeans	RF model	Impact of climate extremes on global agricultural yield	• Temperature-related extremes demonstrated stronger association with yield anomalies • Based on crop yield dataset across ~13 500 spatial units worldwide (1961–2008) • Climatic Research Unit (CRU) TS 3.23 dataset HadEX2 extremes indicator dataset was utilized for climate extremes	(Vogel et al., 2019)
Wheat	CP-ANNs, XY-Fs and SKNs	Yield prediction	• Based on on-line multi-layer soil data, and satellite imagery crop growth characteristics.	(Pantazi et al., 2016)
	SVR, RF, ANN	Yield and protein content prediction	• Based on UAV spectral images and plant heights.	(Zhou et al., 2021)
Rice	SVM, RF	Predicting disease resistance	• Early detection of Rice Sheath Blight Using Spectral Profiles	(Conrad et al., 2020)

			• SVM produced more accurate results	
	ELM	Detection of heavy metal presence	• Based on laser-induced breakdown spectroscopy to detect cadmium content in stems of rice	(Wang et al., 2020)
Maize	GENIE3	GRN detection	• Tissue specific GRN of leaf, root, seed and shoot apical meristem • Tissue-specific GRNs forecast TF regulatory targets	(Huang et al., 2018)
	MLR, SVM, ANN and RF	Above base ground mass study	• Based on structural and spectral information provided by remote sensing from an unmanned aerial vehicle.	(Han et al., 2019)
Tomato	X-means clustering	Fruit characterization	• Based on camera obtained images followed by pixel segmentation, and classification to mature, immature through X-means clustering	(Yamamoto et al., 2014)
	CNN	Pest identification	• To detect Tuta absoluta in tomato plants in early	(Mkonyi et al., 2020)

			stages based on images collected	
Carrot	CNN	Real time classification of weeds	• Based on real time images obtained from field • Faster than conventional CNN	(Knoll et al., 2019)
Sugar beet	R-CNN	Detection of leaf spot disease	• Imaging-based expert (1-3 scale) systems using DL • Updated Faster R-CNN model, developed by changing the parameters of CNN used	(Ozguven et al., 2019)
Soyabean	DL	Root nodule characterization	• Based on root images to determine number and size of nodules on silicon treatment	(Chung et al., 2020)
Potato	FCN	Virus detection	• Real field experiment on hyperspectral images	(Polder et al., 2019)

Rose	PointNet, PointNet++, DGCNN, PointCNN, ShellNet and RIConv	Segmentation of structural parts	• The methods were tested on the ROSE-X data set, containing fully annotated 3D models of real rosebush plants • Best segmentation results were got by PointNet++	(Kaya et al., 2022)
Rapeseed	CNN, MLPN, RBFN	Fungal contamination detection	• Based on the analysis of the morphological structure of rapeseeds was carried out with the use of microscopy	(Przybyl et al., 2020)
Grapes (a vineyard during different seasons)	K-means based classifier	Vegetative index studies[19]	• Based on refined satellite-driven NDVI maps during four different growth seasons	(Mazzia et al., 2020)

Table 2. Various ML models applied in plants to study photosynthetic potential

Plant varieties	ML model used	Sampling data	Phenotype studied	Reference

Wheat	PLSR	Reflectance measurements	Nitrogen per unit leaf area (N_{area}) and leaf dry mass per area (LMA), require laborious, destructive, laboratory-based methods, while physiological traits underpinning photosynthetic capacity, such as maximum Rubisco activity normalized to 25 °C (V_{cmax25}) and electron transport rate (J), require time-consuming gas exchange measurements	(Furbank et al., 2021)
Three lowland seasonal moist tropical forests, including two crane sites in the Republic of	PLSR	Leaf spectroscopic data	To predict V_{cmax}	(Wu et al., 2019)

Panama and one site in Brazi				
Citrus under water stress	RF, SVM, GDboost, Adaboost	Gas exchange and leaf hyperspectral reflectance data	CO2 assimilation rate (Pn), stomatal conductance (Cond) and transpiration rate (Trmmol)	(Zhou et al., 2021)
Lettuce under water stress and bacterial infection	ANN	Reflectance and absorbance measurements	Chlorophyll content	(Osco et al., 2019)
Maize	PLSR and SVR	Hyperspectral reflectance data in the visible, near infrared and shortwave infrared range (VIS–NIR–SWIR, 400–2500 nm)	Chlorophyll content (CHL), leaf water content (LWC), specifc leaf area (SLA), nitrogen (N), phosphorus (P), and potassium (K).	(Ge et al., 2019)

Maize	PLSR and LASSO	Hyperspectral reflectance	Rate of CO_2 assimilation, Transpiration stomatal conductance, intercellular CO_2 concentration, instantaneous water use efficiency, intrinsic water use efficiency, leaf temperature, Chlorophyll, leaf water potential, leaf osmotic potential, leaf osmotic potential at full turgor	(Cotrozzi et al., 2020)
Maize with diverse genotypes, growth stages, treatments with nitrogen fertilizers, and ozone stresses in three growing seasons.	Radiative transfer models (RTMs), data-driven partial least squares regression (PLSR), and generalized PLSR (gPLSR) models	Hyperspectral reflectance	Chlorophyll and Nitrogen content V_{max} prediction	(Wang et al., 2020)

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

- Improved photosynthetic activity enhances the crop productivity
- Machine learning should be used effectively in photosynthesis research
- There is a vast potential for application of ML in photosynthetic pigment studies