Title

An AI-based and coding-free protocol for forests Leaf Area Index (LAI) calculation

Author List

Tao Ma[1,2], Minxue Tang[3], Fangni Liu[1], Akwasi Duah-Gyamfi[4], Stephen Adu-Bredu[4], Mark Arcebal K. Naive[5,6], Imma Oliveras Menor[2,7], Sam Moore[2], Zhiyuan Zhang[8], Yadvinder Malhi[2,9]✉, Cecilia A. L. Dahlsjö[2,9], Huanyuan Zhang-Zheng[2,9]✉

✉ Corresponding Authors

(yadvinder.malhi@ouce.ox.ac.uk)

(huanyuan.zhang@ouce.ox.ac.uk)

Affiliations

1 College of Resources and Environmental Sciences, China Agricultural University, Beijing, China

2 Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

3 Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, United Kingdom

4 Forestry Research Institute of Ghana, Council for Scientific and Industrial Research, Kumasi, Ghana

5 Center for Integrative Conservation and Yunnan Key Laboratory for Conservation of Tropical Rainforests & Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China

6 University of Chinese Academy of Sciences, Beijing 100049, China

7 AMAP – botAnique et Modélisation de l'Architecture des Plantes et des Végétations, Université de Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier CEDEX 5, France

8 College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China

9 Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, United Kingdom

Abstract

Measurements of Leaf Area Index (LAI) have become an important component in modern forest ecology and conservation. Detecting subtle annual, seasonal, and spatial variations in tropical forests is particularly challenging, due to their complex structure and dynamic environmental conditions. Traditional processing of hemispherical photographs involves time-consuming pixel classification that is prone to subjective systematic errors, impeding comparison across months, years, and study sites. While automated methods that ensure consistency are available, they typically require coding expertise, which limits their wider adoption among non-specialists. We present a streamlined protocol that enables automatic, coding-free calculation of LAI by integrating two free software packages: ilastik, for AI-based pixel classification, and CAN-EYE, for LAI estimation.

To evaluate its performance, we compared the seasonal and spatial variations in LAI derived from our protocol with those obtained using traditional pixel classification. Four sub-types of tropical forests were used to test the protocols in different scenarios, including moist evergreen forest, semi-deciduous forest, dry forests, and woody savanna. We found that traditional pixel classification (which requires manual threshold adjustments by operators) was highly influenced by subjective bias, with results varying depending on the time invested in pixel classification. In contrast, the fully automated workflow (our protocol: ilastik pixel classification + CAN-EYE calculation) eliminates subjective bias, enabling consistent cross-years and cross-operator comparisons. Careful validation using field photos and local forest phenology showed that the automated approach captured seasonal LAI variations more accurately than traditional processing, aligning more closely with observed phenological patterns. However, analyses of spatial variation found no significant difference between

protocols, as both captured site-to-site differences with comparable accuracy. Additionally, ilastik can classify pixels into non-binary categories (sky, leaves, and twigs), providing an added layer of detail that is useful for delineating savanna tree phenology. We present an automated, user friendly, and widely applicable protocol that enhances the accuracy, consistency, and reproducibility of LAI measurements, offering a robust tool for advancing tropical forest monitoring and research.

Key words: Leaf Area Index (LAI); digital hemispherical photography (DHP); machine-learning segmentation; ilastik; CAN-EYE; automated workflow; seasonal dynamics

Introduction

Leaf Area Index (LAI) is widely used in climate change models, global carbon cycle assessments, and ecosystem restoration studies (Hazarika et al., 2005; Sumida et al., 2018; Veryard et al., 2023). Accurate measurements of LAI not only contributes to understanding of forests phenology, but also provides essential data supporting carbon cycle modeling (Li et al., 2019; Wang et al., 2024) laying the foundation for dynamically monitoring the response of complex ecosystems to climate change (Wu et al., 2024). Particularly in the context of global climate change, accurately capturing the spatiotemporal variations of LAI (Jiang et al., 2017) has become an essential research requirement in vegetation ecology and environmental science (Houborg et al., 2016; Ma et al., 2023). In tropical forests, LAI exhibits subtle but important spatial and temporal variation (de Wasseige et al., 2003; Malhado et al., 2009). Such variation is smaller than global scale variation but critically important for understanding the functioning of tropical forests. Tropical forests contain several subtypes including wet evergreen, semi-deciduous, dry forests, and more. Seasonality in wet evergreen tropical forests is not apparent to human eye, but such minor seasonal variations reveal the sensitivity of forest phenology to changes in climatic factors (Wu et al., 2017, 2016).

Accurately detecting these small fluctuations in LAI has always been a significant challenge (Li et al., 2024; Yan et al., 2019; Zou et al., 2023). Global products, such as the widely used MODIS product, are often critiqued after comparisons with field measurements (Zhang-Zheng et al., 2024b). Although drone-based remote sensing methods (Yan et al., 2024, 2021) provide large-scale LAI estimates, they typically rely on expensive LiDAR sensors or require powerful computing resources, creating significant financial barriers (Gara et al., 2021). Moreover, many remote sensing algorithms and models lack transparency in data and code, leading to poor reproducibility and exacerbating global disparities in research applications (Kedron and Frazier, 2022). Professional LAI measurement devices, for example Licor-LAI2200C, provide consistency but their prohibitive cost significantly limits widespread usage across resource-constrained institutes. Consequently, Digital hemispherical photography (DHP) methods through digital single-lens reflex (DSLR) cameras remain the primary ground-based approaches for quantifying forest LAI, especially in tropical regions. However, this approach faces a major challenge, among many others: each DHP image needs to be manually processed for sky-leaves pixel classification. This step is very time-consuming and incurs strong

subjective segmentation errors which affect consistency in cross-ecosystem studies (Fang et al., 2014).

The DSLR DHP method uses fisheye lenses to capture canopy images and, based on the segmentation of leaves and sky, calculates a 'gap fraction' to indirectly estimate LAI (Yang et al., 2023). Many algorithms have been proposed for this calculation, which introduce systematic errors to the absolute value of LAI, but not the pattern in seasonal or inter-annual variation. To delineate temporal (or spatial) pattern of LAI, it is crucial to ensure consistency in the segmentation step (i.e. the gap fraction step). However, subjective errors in the segmentation step impede cross-study comparisons (Jonckheere et al., 2004). Considerable variations exist among researchers in arbitrating the threshold for leaf-sky segmentation, leading to difficulties in standardising data across experiments and even across different years at the same site (due to change of personnel) (Pyare and Henning, 2012). Additionally, traditional DHP data analysis software (e.g., CAN-EYE) relies heavily on user input, requiring manual marking and parameter adjustments, which further exacerbates inconsistency (Gilardelli et al., 2018). Automatic image processing methods have been proposed. For instance, a MATLAB-based script was introduced to automate gap fraction analysis of digital canopy images (Fuentes et al., 2008). A modified image segmentation approach incorporating color space transformations and thresholding was developed to exclude non-leaf elements in fruit tree canopies (Carrasco-Benavides et al., 2016). These methods provide automation and thus theoretically avoid subjective bias, but the complexity of mathematical formulas and codes make them difficult to use for researchers without expertise in computer science (Chen et al., 2022). Consequently, these methods often lack reproducibility (Baker, 2016; Hoyt and Gyori, 2024) further exacerbating inconsistencies. There is therefore an urgent need to summarise the decades of LAI practice and establish a protocol that is free of subject errors, low-cost and coding free so that it can be widely applied for better standardisation of LAI measurements globally, and for decadal tracking of LAI.

In this paper, we are providing a protocol that leverages free, user-friendly software (ilastik and CAN-EYE) to minimise subjectivity in image segmentation. This protocol is specifically designed for a diverse user base, including forestry practitioners as well as undergraduate and graduate students seeking for cost-effective, reproducible LAI measurements with enhanced sensitivity to detect subtle LAI variation. The protocol has been applied in several forests ecological studies which manage to delineate subtle seasonal, spatial and annual variations in forest LAI. To validate the protocol we will also conduct a comparison between the protocol

and the traditional manual pixel classification method. Through comprehensive analysis of seasonal and spatial variation patterns across three representative tropical forest types (moist evergreen, semi-deciduous, and woody savanna), we show key operational limitations in manual methods while demonstrating the AI-driven automation significantly enhances measurement consistency.

Methods

Case study site

West African forests exhibit a well-defined ecological gradient, transitioning from coastal wet evergreen forests to savanna woodlands as they approach arid regions. To capture this diversity, our study sites encompass four distinct forest types: 1) wet evergreen, 2) semi-deciduous, 3) dry forests, and 4) woody savanna. The Ghanaian forests examined in this research are located at the eastern edge of the Upper Guinea forest, a biogeographic region separated from the Lower Guinea forest by the notable Dahomey Gap (Luiselli and Fa, 2019; Prance, 1984). West African forests show a pronounced ecological gradient from coastal wet evergreen forests to savannas with the latter exhibiting stronger rainfall seasonality than the former (Fig. 1). Our study focuses on three reserves along this gradient in Ghana (Table 1). Overall, there are 14 one-hectare plots, classified into four forests types and located in three natural reserves.

Study sites from part of the Global Ecosystem Monitoring (GEM) network (Malhi et al., 2021) and are registered in the ForestPlots.net database. Each monitoring plot (1-ha) was spaced at intervals ≥100 km between sites, providing a network for multi-scale LAI comparisons (both between-site vs. within-site variability).

Data Collection and processing

In our case studies 25 hemispherical photo sampling points systematically distributed within each one-hectare plot. Note that the protocol mentions that the number of sampling points could be much less than 25. Details in Methods are describing the case studies. For the protocol please check the link in supplementary. Photos were taken monthly from March 2012 to July 2019 for the study site BOB, while ANK and KOG initiated monitoring later. The dataset comprised,

19,044 hemispherical photos, demonstrating the methodological protocol's capacity for consistent processing of large image datasets. Following the protocol, this workflow generated 711 LAI values, as the CANEYE software outputs one value per plot-month combination. Spatial, seasonal, and interannual variations were analysed by averaging these values across relevant temporal scales. For interannual variation analysis, only years with more than four months of data were retained. To facilitate interpretation of seasonal LAI variation, representative field photographs and corresponding hemispherical images are shown (Fig. 2).

Across the study sites from ANK, BOB to KOG, a gradient of increasing rainfall seasonality was observed (Fig.1). The Seasonality Index (SI) was computed following this formula:

$$SI = (1/R) * \Sigma|Xj - R/12|$$

where Xj is monthly precipitation and R is total annual rainfall (Imteaz and Hossain, 2023).

We processed CHIRPS daily precipitation datasets (2000–2020) using Google Earth Engine(Funk et al., 2015), where we aggregated rainfall data to monthly totals, computed annual rainfall sums for each pixel, and derived the SI values for each year. The final output represents the 21-year mean SI, which characterizes long-term rainfall seasonality patterns across the study area.

Table 1. Key characteristics of study sites

Study region	ANKASA (Wet)	BOBIRI	KOGYAE *	KOGYAE *
Latitude (°)	5.2680	6.6910	7.2616	7.3012
Longitude (°)	-2.6955	-1.3389	-1.1501	-1.1649
Elevation (m asl)	114	235	229	221
Forest Type	Evergreen forest (EF)	Semi-deciduous forest (SDF)	Dry forest (DF)	Woody savanna (WS)
Mean Annual Precipitation	2050 mm	1500 mm	1200 mm	1200 mm

Plot Details	3 × 1-ha plots (ANK-01–03)	6 × 1-ha plots (BOB-01–06)	2 × 1-ha plots (KOG-02–03)	2 × 1-ha plots (KOG-05–06)

*KOG-04 is a forest-savanna transition plot, it is included in the results but it does not belong to any of the four forest types.

The protocol

The protocol provides a comprehensive, step-by-step guide for three key processes: 1) image segmentation (using ilastik), 2) LAI calculation (using CANEYE) and 3) LAI uncertainty quantification. To enhance usability, the protocol includes extensive annotated screenshots and instructional videos (Supplementary video). All software parameters for both ilastik and CANEYE are documented in detail within (Supplementary protocol). The protocol's strength include jargon-free explanations and minimal use of mathematical equations or programming code. Furthermore, it is compatible with low-tier commercial laptops without GPU acceleration. Ilastik is an open-source image analysis software that applies machine learning techniques (Berg et al., 2019). We selected both Ilastik and CANEYE because they are freely available and do not require programming expertise. However, users have the flexibility to combine ilastik outputs with alternative LAI estimation software packages (e.g., 'hemispheR', 'WinSCANOPY', 'hemisfer'). For comparative purposes, we present results from both our protocol (Approach 1) and the traditional CANEYE approach (Approach 2).

Approach 1 The protocol workflow.

In the protocol, we used ilastik to process fisheye photos which classified canopy and sky pixels into black and white. Note that CAN-EYE also has a function to classify pixels which is used in Approach 2 not Approach 1.

For the case study, we developed two distinct ilastik classification models: 1) a model trained on sites ANK and BOB (forests), and 2) a separate model for the KOG site (dry forests and woody savanna). The methodological distinction was necessary due to fundamental differences in phenology: one of the KOG plots exhibited complete winter defoliation during the dry

season (February), requiring a three-class annotation system (leaf, sky, branches) compared with the standard binary classification applied at ANK and BOB.

For model training, we randomly selected twenty fisheye photographs from each of the ANK and BOB sites. Pixel annotation was performed using the Pixel Classification workflow in ilastik, with manual labelling of leaf and sky pixels. The software then learned from these annotations to generate a classification model, which was applied in batch mode to process all fisheye images across all plots and months at both ANK and BOB sites. An identical workflow was implemented for KOG, with the modification of using the expanded three-category annotation scheme. Further details of the image processing steps are provided in the protocol section.

The black-and-white images classified by ilastik were organized by plot and month before being processed through the Binary DHP interface in CAN-EYE. It should be noted that CANEYE calculates a single LAI value per plot-month combination as standard output due to its algorithm (25 images per calculation in this study). CANEYE could not provide LAI for individual image.

Hereafter, results from the above processing are referred to as 'the protocol'.

Approach 2: Leaf Area Index Measurement Based on CANEYE Color Image Processing interface following the Traditional workflow

In approach 2, we avoid Ilastik and use the pixel classification function embedded in CANEYE. The colored raw fisheye images were organized into plot-month folders and processed manually through CAN-EYE. Each folder contained 25 images. LAI values were calculated using the Color DHP (RGB) interface in CAN-EYE. Unlike approach 1, this method involved manual processing of a subset of plot-month folders (detailed below) due to the impracticality of processing all 19,044 images. The time-consuming nature is one of the drawbacks of approach 2. In approach 1, 40 images served as training data for developing two models (20 images per model), which were then applied to all 19,044 images. By contrast, approach 2 required manual processing of each individual image.

To demonstrate the impact of subjective errors on CANEYE results, we designed a controlled experiment incorporating three distinct time allocations for the pixel classification process: 30 seconds, 2 minutes, and ‘unlimited time’ per plot-month package. Note that in Approach 1, we spent much longer time (even longer than ‘unlimited time’) on each image to make the outcome almost perfect, simply because there we only have 40 images to process.

For all plot-month combinations in the ANK, BOB, and KOG forests post-January 2018, we implemented Color DHP processing where the pixel classification step was strictly limited to 2 minutes per plot-month package. During processing, we mainly marks 2-3 representative photographs from each 25-photo set for manual classification of sky and canopy elements. This experiment mimic typical CANEYE users scenarios.

For the ANK-01 and BOB-01 study plots post-January 2018, we implemented the following Color DHP processing (designated as ‘unlimited time’). This protocol imposes no restrictions on either the quantity of selected photos or the number of manual classifications performed per image. The process concludes only when the operator achieves full satisfaction with the sky-canopy classification accuracy. This experimental design replicates scenarios involving meticulous, perfectionist users, ensuring the highest possible classification quality.

For the ANK-01 and BOB-01 study plots post-January 2018, we implemented the following Color DHP processing (designated as ’30 seconds’). During the process, we mainly mark 1 photo from the 25 photos for manual classification, and this photo is manually classified 3-4 times on average. The classification takes exactly 30 seconds. This experiment mimic hasty users scenarios.

In this study, the operators conducting the CANEYE image segmentation experiment are not responsible for calculating the Leaf Area Index (LAI). During segmentation, they perform only manual sky-canopy classification without access to LAI-related results. At this stage, operators view solely the sky percentage data and do not receive the final LAI value. The LAI computation occurs post-segmentation, automatically by the software, ensuring the classification phase remains independent of LAI derivation. Consequently, the segmentation operators remain free from subjective bias, as their work in uninfluenced by LAI outcomes.

Hereafter, results from the above processing are referred to as ‘30 seconds’, ‘2 minutes’ and ‘unlimited time’.

Results

Seasonal variation of LAI

The study was conducted at three forest sites in Ghana, West Africa, each situated in areas with different rainfall seasonality (Fig.1), which resulted in differences in phenology. To aid in the interpretation of seasonal LAI variations, representative field and hemispherical photographs are shown (Fig. 2).

For the moist evergreen forest ANK-01, LAI of 'the protocol' exhibits clear seasonal variation. From April to July 2018 and March to July 2019, the LAI shows a noticeable seasonal increase, reaching its peak in July 2018 and July 2019, respectively. Between August and December 2018, the LAI remains in a fluctuating phase. In contrast, during January to April 2018 and December 2018 to February 2019, the LAI values are relatively lower (Fig. 3). These seasonal variations are consistent with the multi-year average for the study sites (Supplementary Figure 1).

In traditional CAN-EYE color image processing (approach 2), it is important to note that the results vary depending on the time spent on the operation. The suspected reason is that operators often mark up the sky first and then spend most of their time marking leaves and branches while iteratively checking the segmented images. Thus, the '30 seconds' experiment likely missed some leaves due to time constraints. Ilastik does not have an 'operation time' issue since the random forest algorithm is highly efficient. It achieves excellent pixel classification with very few manual marks. Moreover, the training dataset for Ilastik is very small (ca. 20 photos), allowing the operator review every image to ensure precise segmentation. Because each month is processed independently in traditional CAN-EYE, there is more month-to-month variation compared to 'the protocol', which is less desirable since the seasonal variation of forest LAI should be smooth from a physiological perspective.

'30 seconds' performed poorly in terms of both seasonal variation and underestimated LAI values (Fig. 3). In contrast, '2 minutes' exhibits seasonal trend highly similar to those of 'the protocol'. Nonetheless, statistical analysis confirmed that the LAI values from the '2 minutes' are significantly higher than those obtained with 'the protocol' (Supplementary Figure 11 and Figure 13).

In the CAN-EYE traditional processing, both '2 minutes' and 'unlimited time' found exceptionally high LAI values in September 2018. However, based on visual inspection of the color photographs (Supplementary Figure 2), it is more likely that July 2018 should have had the highest LAI. Additionally, such considerable fluctuations in LAI from July (high value), to August (low value) and back to September (very high value) are biologically implausible. In summary, our detailed seasonal comparison at the ANK01 site demonstrates that 'the protocol' provides more realistic results for seasonal variation.

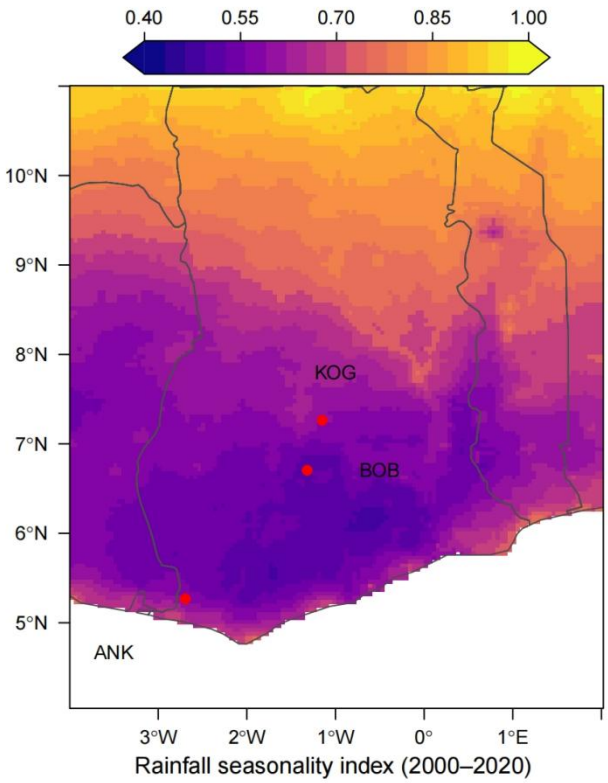


Fig. 1 Map of the study site and rainfall seasonality index of Ghana, West Africa *. The background shows the seasonality index of rainfall. Typically, values smaller than 0.2 suggest no seasonality. Values around 0.45 suggest moderate seasonality with a short drier season. Values larger than 1.0 suggest most rainfall occur in less than 3 months* (Nair et al., 2014).

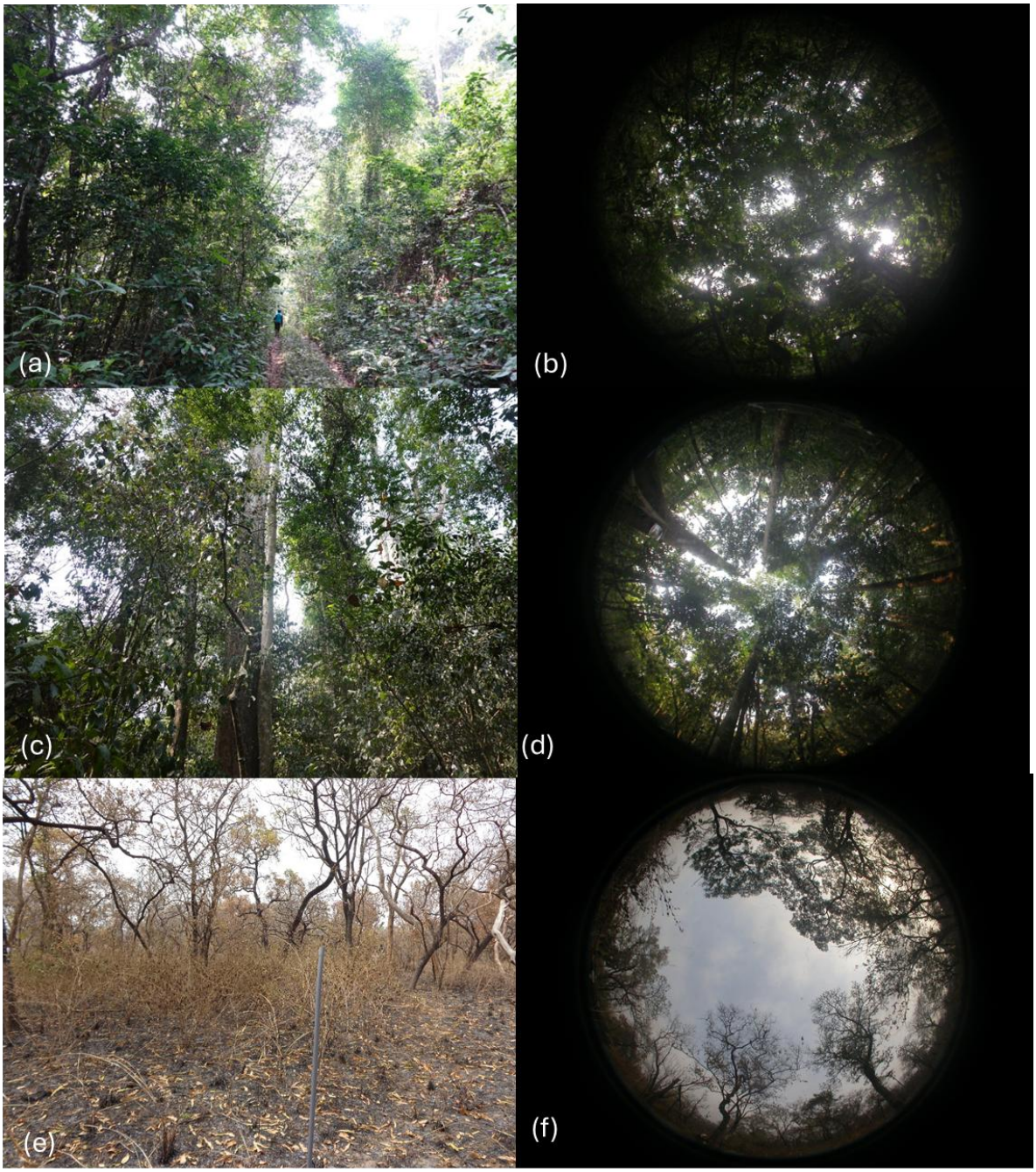


Fig. 2 Photos of the study sites and examples of their hemispherical photos. *Here we show (a) the path leading to site ANK, at the entrance of ANK; (b) hemispherical photo of plot ANK01 in January 2019 (dry season); (c) study plot BOB-01 (d) hemispherical photo of BOB-02 in January 2019; (e) Study plot KOG-05 after fire (f) hemispherical photo of KOG-05 taken in 2019 January.*

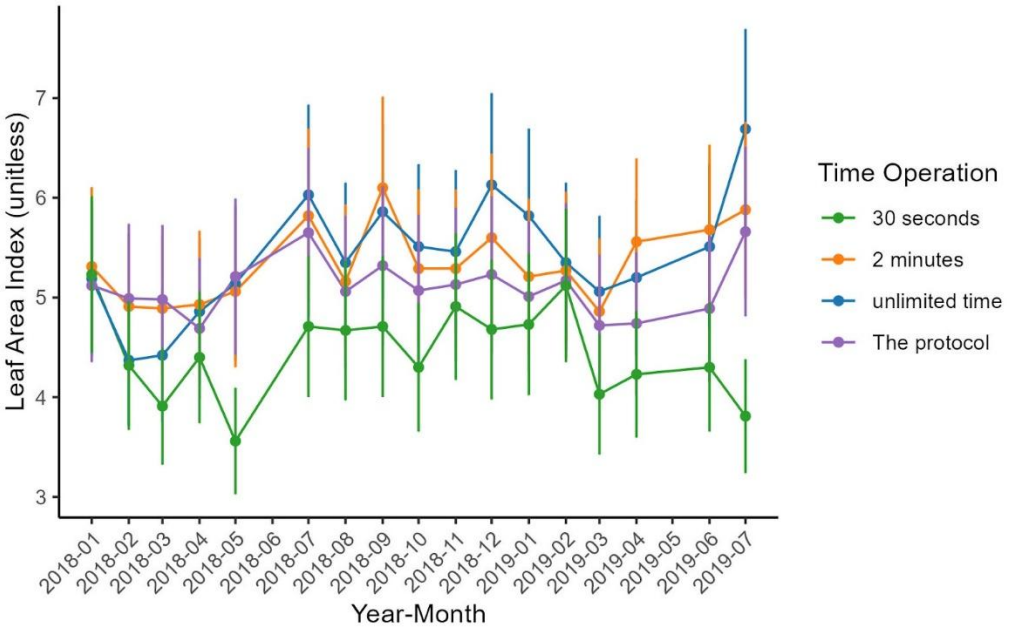


Fig.3 Time series of Leaf area index for different experiment settings in the Ankasa National Forest (Moist Evergreen Forest) plot ANK-01*. The purple line represents the protocol, using ilastik to process images, which has no operation time as it only needs to be trained once. The remaining colours represent traditional manual image processing experiments (Approach 2). Error bars represent ± 15% of the mean.*

For the BOB-01 forest, both 'the protocol' and '2 minutes' methods showed very similar seasonal variations. It is particularly evident that from November 2018 to April 2019, the LAI values remained at very low levels. This result aligns well with the dry season months in southern Ghana (the dry season lasts from December to mid-March).

Statistical analysis shows that the LAI of '2 minutes' are consistently and significantly higher than those of 'the protocol' in most months (Supplementary Figure 11 and Figure 13), mirroring the pattern observed at ANK-01. In the careful processing (blue line), although the seasonal variation remains consistent, the LAI values are considerably higher than those from both 'the protocol' and '2 minutes'. The '30 seconds' operation (green line) produces LAI values considerably lower than those from 'the protocol' and '2 minutes'. (Fig. 4). Similar to the conclusion for ANK-01, the results are subject to operation time, and the traditional CAN-EYE method has more monthly fluctuation.

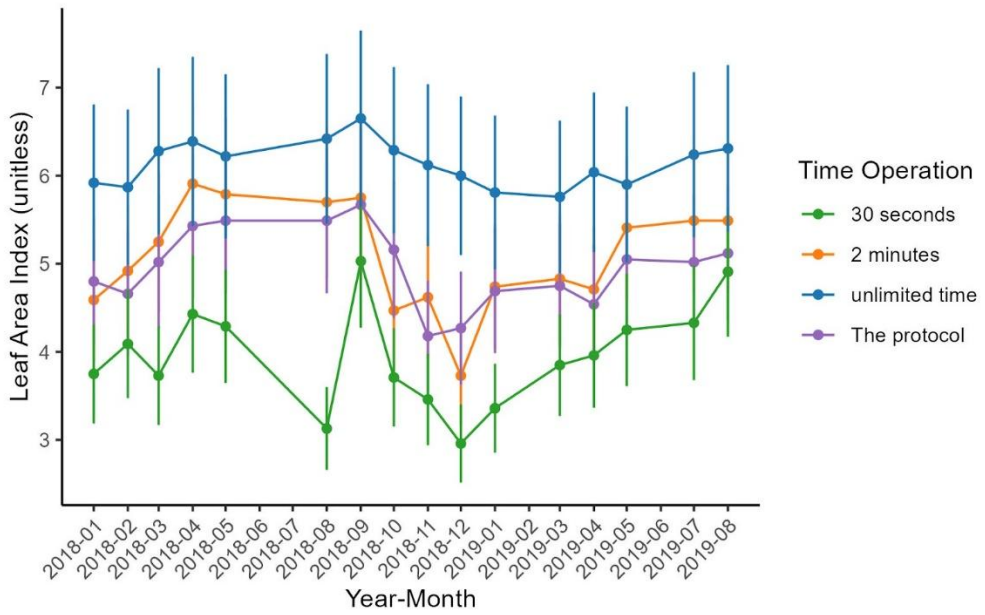
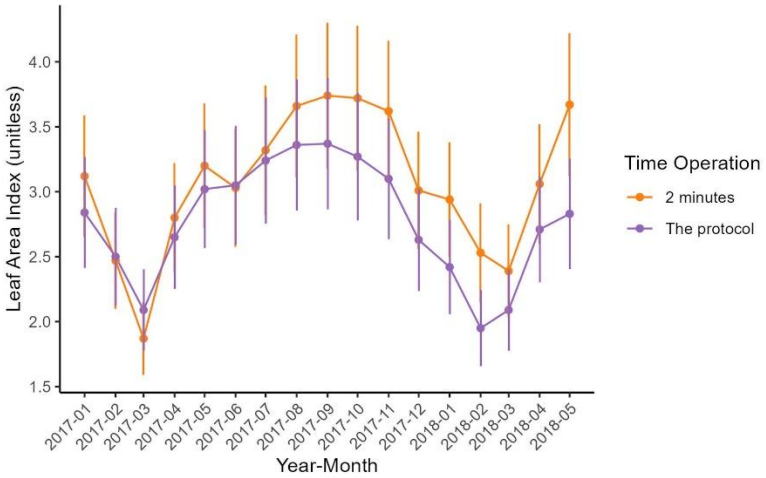
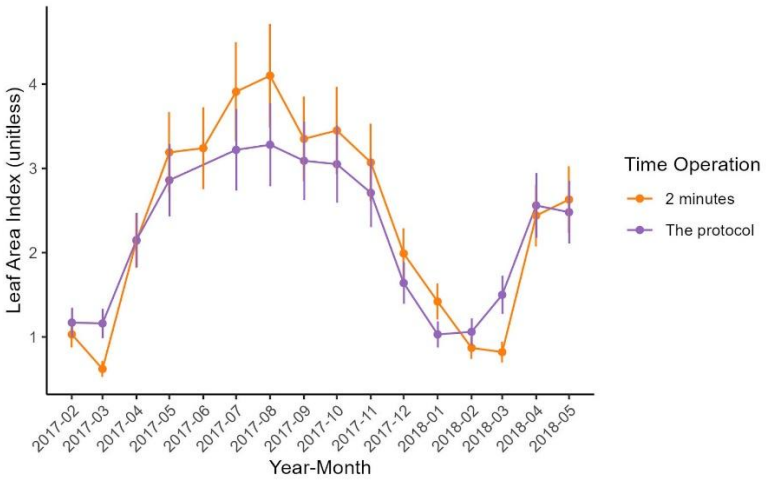


Fig.4 Time series of Leaf area index (LAI) for different experiment settings in the Bobiri (semi-deciduous Forest) plot BOB-01. *The purple line represents the protocol, using ilastik to process images, which has no operation time as it only needs to be trained once. The remaining colours represent traditional manual image processing experiments (Approach 2). Error bars represent ± 15% of the mean.*

For the KOG forest, both 'the protocol' and '2 minutes' showed similar seasonal variation. LAI of '2 minutes' is significantly higher than that of 'the protocol' at KOG-02 and KOG-03. No significant difference was found for other KOG plots (Supplementary Figure 12). At KOG, some of the photo sampling points feature a large sky window. Hence, each month-plot package comprises pictures with and without large sky windows. CAN-EYE set a constant threshold for the whole package which often fails to process the two distinct types of photos (Supplementary figure 5). When the marking is drawn for closed canopy photos, a portion of the large sky window is often mistaken as leaves, leading to over-estimation of LAI. In contrast, ilastik avoids this dilemma entirely, significantly reducing processing time.



(a)



(b)

Fig.5 Time series of Leaf area index (LAI) for different experiment settings in the Kogyae (dry forest). *The purple line represents the protocol which uses ilastik to process images. It has no operation time as it only needs to be trained once. The orange line represents the two-minute processing experiment (Approach 2).* ***(a)*** *KOG-02 forest(dry forests)；* ***(b)*** *KOG-06 forest(woody savanna). Error bars represent ± 15% of the mean.*

Spatial and inter-annual variation of LAI

When compared to the ANK and BOB forests, the KOG forest sites exhibits a lower LAI, consistent with its ecotone position between dry forests and savanna. For 'the protocol', the highest LAI value occurs at ANK-03, whereas for '2 minutes', the highest LAI is found at the BOB-04 site (Fig. 6). The difference between BOB-04 and ANK-03 is difficult to discern through visual inspection of the color photographs (Supplementary Figure 6), and these two

plots have very similar stem biomass (Zhang-Zheng et al., 2024a). Additionally, statistical analysis confirmed that the LAI values obtained from 'two minutes' are consistently—and significantly—higher than those derived from 'the protocol', as discussed above (Supplementary Figure 12). Overall, both 'the protocol' and 'two minutes' demonstrate a reasonable pattern of LAI variations between plots (Fig. 6), which aligns with the biomass variation of these plots (Zhang-Zheng et al., 2024a). For inter-annual variation, a comparison between '2 minutes' and 'the protocol' was not possible due to insufficient manpower to to manually process 19,044 images. Nonetheless, using 'the protocol', we demonstrated that ANK and BOB experienced an increase in LAI over time, while KOG showed a decrease (Fig. 7). Forest plots within each site are separated by only several kilometers, so we would expect to see similar inter-annual variation in LAI. As expected, the inter-annual pattern is consistent among plots within each study site. For BOB and KOG, we found that LAI declined in 2016, consistent with the 2015/2016 El Nino event.

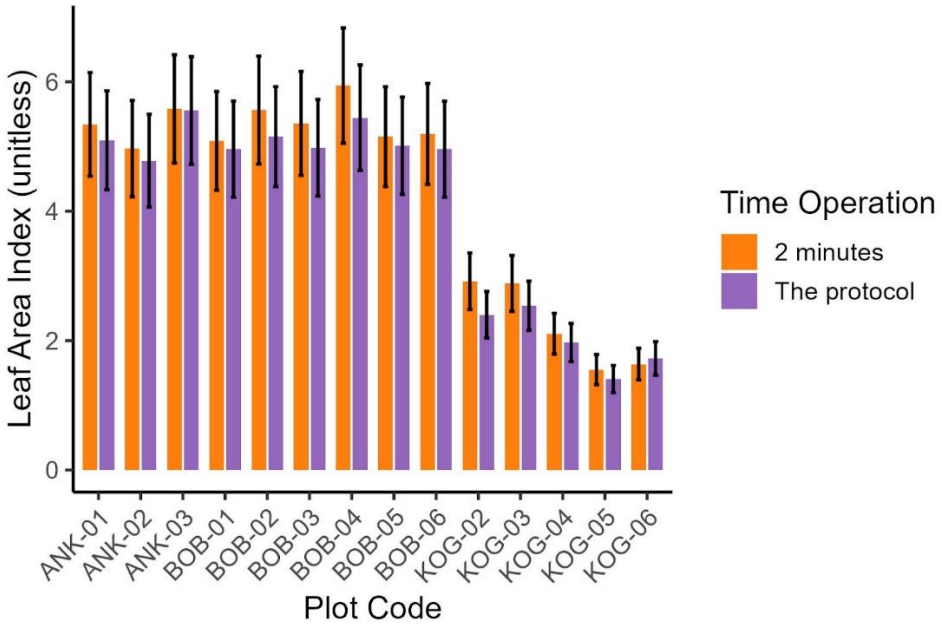


Fig.6 Comparison of LAI between 2-minutes processing experiment and the protocol (fully automated using ilastik) across different forest plots. *The orange bars represent CAN-EYE 2-minute processing, while the purple bars represent ilastik fully automated processing. The Plot Code includes ANK (Moist Evergreen Forest), BOB (Semi-Deciduous Forest), KOG 02-03 (Dry Forest) and KOG 05-06 (Savanna). Error bars represent ± 15% of the mean.*

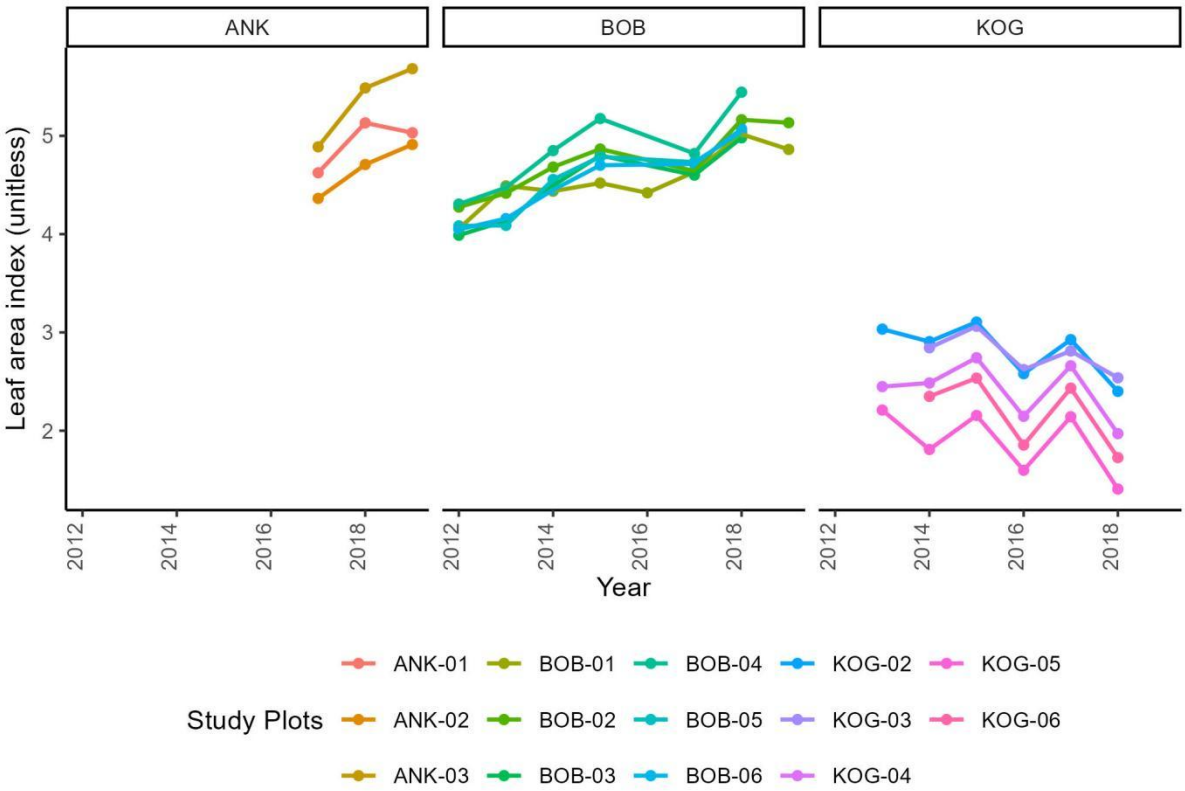


Fig.7 Inter-Annual Variation Of Leaf Area Index Following The Protocol. *Different colours represent study plots.*

Discussion and conclusion

The traditional manual pixel classification

This study reveals that the traditional manual pixel classification process introduces a strong systematic bias in LAI estimates. Variations among different operators arise when processing the same image, depending on the time spent on annotation, their experience, and subjective judgment criteria.

First, a 'grey zone' exists at the edge of leaves (Supplementary Figure 7) where human eye could not identify whether the pixels belong to leaves or sky. When performing manual

classification, operators must arbitrate for these grey zones. However, sun-bleached leaf edges or light green areas are often misclassified as sky, which reduces the reliability of the data. Previous studies have demonstrated that human judgment in complex scenarios significantly impacts data results (Jonckheere et al., 2004). Consequently, more grey zones are identified as leaves when operators spend more time on leaf annotation, resulting in higher LAI values with increased working hours (Figure 4). For Ilastik, the AI also needs to find a middle ground for these grey zones. The AI's decision may not be precise, but it is consistent across years and sites.

Furthermore, field photographs often exhibit varying exposure levels. If the lighting is dimmer, operators may subjectively interpret the foliage as denser, while variations in exposure can affect the size of the grey zones. Under lower light conditions, the distinction between leaves and sky become more pronounced.

In addition, CAN-EYE's edge processing demonstrates a lack of precision, frequently introducing erroneous pixel clusters along sky boundaries (Supplementary Figure 8), likely due to a programming artifact. These flaws directly contribute to systematic biases in true PAI, resulting in a consistent overestimation of LAI. Previous studies have documented similar classification errors in LAI measurements (Duveiller and Defourny, 2010; Zhu et al., 2023).

Secondly, CAN-EYE relies on manually set thresholds (such as the gamma factor), to optimize classification results. However, this process lacks a standardization. Studies have demonstrated that threshold variations can cause gap fraction deviations of up to 17% (Xie et al., 2023). This methodological inconsistency limits the tool's applicability for multi-site and long-term monitoring studies.

By contrast, ilastik utilizes a random forest model which could be automatically applied to all photos, thereby significantly mitigating the subjectivity issues associated with manual classification. Another key strength is its ability to categorize image pixels into three classes: sky, leaves, and twigs. Unlike traditional binary classification methods, where leafless canopy (but full of twigs) is often misclassified as leaves leading to LAI overestimation, particularly in savanna ecosystems (Supplementary Figure 9). Ilastik's ternary annotation can identifies twigs beyond leaves and sky. This approach effectively eliminates the misclassification of leafless canopy as leaves, thereby substantially improving LAI measurement accuracy in woody savanna or dry forests in the dry season, when part of the tree community shed all leaves.

Note that this is different to previous studies on identifying stems and branches of conifer using multispectral images (Kucharik et al., 1998). Ilastik still recognizes main stems as leaves.

Efficiency and accuracy

Ilastik demonstrates significant efficiency advantages. With CAN-EYE, processing images for a given plot-month typically requires between 2 to 20 minutes, depending on image complexity, operator experience, and computer performance. By contrast, ilastik requires training on only approximately 20 representative images to generate a high-quality model, enabling fully automated, unmanned processing thereafter. This efficiency has been further validated in other ecological data processing applications (Horning et al., 2020), highlighting its potential for broader scientific use. Moreover, the seasonal variation of LAI derived from ilastik closely tracks rainfall patterns, and its spatial distributions align consistently with expected aboveground biomass differences across sites.

A low cost, coding free, and wide applicable protocols for research equity

As demonstrated above, using an AI tool for leaves-sky segmentation can considerably reduce subjective errors and improve consistency, both of which are critically important for delineating subtle spatial and temporal variations in tropical forests LAI. In fact, many previous studies have proposed new AI methods for LAI calculation, demonstrating higher accuracy and efficiency than traditional methods (Yamaguchi et al., 2020). However, tropical forests researchers (particularly those in developing countries where most tropical forests are located) often lack high-performance computing resources and coding expertise, which limits the application of complex AI methods (Ha, 2024; Sarkar and Chapman, 2021). To facilitate comparison across time and study sites, there is an urgent need for a widely applicable AI-based LAI protocol.

Against this backdrop, our protocol offers a low-cost, flexible, and user-friendly solution for applying AI to LAI calculation, opening new avenues for research teams with limited resources (Kreshuk and Zhang, 2019). Since both software programs in the protocol are free, the only required purchase is a digital camera with a fisheye lens, resulting in a total cost of less than £1,000 (Zhu et al., 2023). Using a smartphone with a fisheye lens is also possible, provided that the appropriate parameters are used in CAN-EYE. Moreover, ilastik uses established, classic AI algorithms that can be run on a standard commercial laptop (around £600), neither a high-performance cluster nor an advanced GPU in required (C. Sommer et al., 2011).

As we have shown, ilastik's random forest models can be applied to various forest types, including woody savanna, as the software allows personalized training based on the characteristics of different ecosystems (Obels, 2021). Thus, the protocol can be widely applied to enable site-to-site comparison while being free of subjective bias. Moreover, the protocol does not regulate the DHP images-shooting process, meaning it could be applied to photos that were already taken in the past.

The protocol's low cost and high versatility have profound implications for global ecological research equity. Research teams in developing countries can leverage this approach to achieve precise LAI measurements with limited equipment budgets, thereby promoting greater democratization of regional ecological studies. This automated, reproducible workflow is equally suitable for research projects across various scales while also laying the solid foundation for international ecological research collaboration.

It should be noted that this comparative analysis focuses on CAN-EYE, a widely used LAI calculation software, and does not consider other LAI software such as Hemisfer. Nonetheless, ilastik could be combined with Hemisfer and other LAI calculation software. We chose CAN-EYE rather than Hemisfer because the latter is not free of charge. Hemisfer itself can process images automatically completely without supervision (called 'batch processing') by 'Automatic thresholding' (XXXX). However, we could not get clear seasonal pattern of LAI out of this function likely due to issues in the black-white threshold. Replacing the 'Automatic thresholding' function with ilastik may improve the outcomes from Hemisfer, but testing and showing results of Hemisfer are outside the scope of this study, which nonetheless is suggested for future researchers.

Conclusion

Many traditional DHP software programs rely on manual pixel classification (distinguishing between leaves and sky), which may introduce subjective errors (Y. Tong et al., 2022), while AI-based automatic methods often require coding expertise. In this study, we introduce a protocol that combines two user-friendly softwares, ilastik and CAN-EYE, to achieve AI-based automated classification and LAI calculation. By comparing this approach with traditional processing procedures, this study demonstrates the importance of using an AI-based software for the leaf-sky pixel classification step.

The protocol offers several key advantages:

(1) High accuracy: By applying the same ilastik model consistently across various study sites and measurement periods, the method reliably captures subtle changes in LAI. This approach can distinguish both interannual and seasonal variations in tropical LAI and can differentiate between tropical forest subtypes.
(2) Efficiency and consistency: Compared to traditional manual methods, this approach significantly reduces both manual processing time and subjective errors. Because the LAI outputs are not operator-dependent, they achieve high reproducibility.
(3) Accessibility and cost-effectiveness: The software package feature intuitive interfaces that require no coding skills, making them accessible to users without programming expertise. Since both software options are free, the camera and tripod represent the only required costs. Consequently, the protocol offers technical support for LAI measurement in lower-income countries, potentially enhancing ecosystem monitoring equity.

Codes and data availability

All data and codes generated in this study have been deposited in the 'zenodo' database under accession code **[XXXX will be uploaded upon acceptance]**. The above includes the protocol. As Zenodo allows authors to archive a new version after publication. We are likely to update

the protocol by uploading a new version on Zenodo. Please check the Zenodo archive for possible new versions. In the protocol, we mentioned that users could use **Image_conversion_20220407.m** and **lets_change_values.R** instead of the 'Renormalize' function of ilastik to replace values in the classification output images. These codes are not essential for people working with CANEYE, but could be useful to combine ilastik with other LAI software. In the protocol, we mentioned that **Gather_LAI_fapar_from_caneye.R** can be used to consolidate output Excel files, rather than manually opening each file.

References

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454. https://doi.org/10.1038/533452a

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., Eren, K., Cervantes, J.I., Xu, B., Beuttenmueller, F., Wolny, A., Zhang, C., Koethe, U., Hamprecht, F.A., Kreshuk, A., 2019. ilastik: Interactive machine learning for (bio)image analysis. Nat. Methods 16, 1226–1232. https://doi.org/10.1038/s41592-019-0582-9

C. Sommer, C. Straehle, U. Köthe, F. A. Hamprecht, 2011. Ilastik: Interactive learning and segmentation toolkit, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Presented at the 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 230–233. https://doi.org/10.1109/ISBI.2011.5872394

Carrasco-Benavides, M., Mora, M., Maldonado, G., Olguín-Cáceres, J., Bennewitz, E. von, Ortega-Farías, S., Gajardo, J., Fuentes, S., 2016. Assessment of an automated digital method to estimate leaf area index (LAI) in cherry trees. N. Z. J. Crop Hortic. Sci.

Chen, Q., Zheng, B., Chenu, K., Hu, P., Chapman, S.C., 2022. Unsupervised plot-scale LAI phenotyping via UAV-based imaging, modelling, and machine learning. Plant Phenomics 2022. https://doi.org/10.34133/2022/9768253

de Wasseige, C., Bastin, D., Defourny, P., 2003. Seasonal variation of tropical forest LAI based on field measurements in central african republic. Agric. For. Meteorol. 119, 181–194. https://doi.org/10.1016/S0168-1923(03)00138-2

Duveiller, G., Defourny, P., 2010. Batch processing of hemispherical photography using object-based image analysis to derive canopy biophysical variables. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 38.

Fang, H., Li, W., Wei, S., Jiang, C., 2014. Seasonal variation of leaf area index (LAI) over paddy rice fields in NE China: Intercomparison of destructive sampling, LAI-2200, digital hemispherical photography (DHP), and AccuPAR methods. Agric. For. Meteorol. 198–199, 126–141. https://doi.org/10.1016/j.agrformet.2014.08.005

Fuentes, S., Palmer, A.R., Taylor, D., Zeppel, M., Whitley, R., Eamus, D., 2008. An automated procedure for estimating the leaf area index (LAI) of woodland ecosystems using digital imagery, MATLAB programming and its application to an examination of the

relationship between remotely sensed and field measurements of LAI. Funct. Plant Biol. 35, 1070. https://doi.org/10.1071/FP08045

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., Michaelsen, J., 2015. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci. Data 2, 150066. https://doi.org/10.1038/sdata.2015.66

Gara, T.W., Rahimzadeh-Bajgiran, P., Darvishzadeh, R., 2021. Forest leaf mass per area (LMA) through the eye of optical remote sensing: A review and future outlook. Remote Sens. 13, 3352. https://doi.org/10.3390/rs13173352

Gilardelli, C., Orlando, F., Movedi, E., Confalonieri, R., 2018. Quantifying the accuracy of digital hemispherical photography for leaf area index estimates on broad-leaved tree species. Sensors 18, 1028. https://doi.org/10.3390/s18041028

Ha, D.H., 2024. The application of artificial intelligence in addressing climate change in tropical regions, in: 2024 International Conference on Platform Technology and Service (PlatCon). Presented at the 2024 International Conference on Platform Technology and Service (PlatCon), pp. 89–94. https://doi.org/10.1109/PlatCon63925.2024.10830746

Hazarika, M.K., Yasuoka, Y., Ito, A., Dye, D., 2005. Estimation of net primary productivity by integrating remote sensing data with an ecosystem model. Remote Sens. Environ. 94, 298–310. https://doi.org/10.1016/j.rse.2004.10.004

Horning, N., Fleishman, E., Ersts, P.J., Fogarty, F.A., Wohlfeil Zillig, M., 2020. Mapping of land cover with open-source software and ultra-high-resolution imagery acquired with unmanned aerial vehicles. Remote Sens. Ecol. Conserv. 6, 487–497. https://doi.org/10.1002/rse2.144

Houborg, R., McCabe, M.F., Gao, F., 2016. A spatio-temporal enhancement method for medium resolution LAI (STEM-LAI). Int. J. Appl. Earth Obs. Geoinformation 47, 15–29. https://doi.org/10.1016/j.jag.2015.11.013

Hoyt, C.T., Gyori, B.M., 2024. The O3 guidelines: Open data, open code, and open infrastructure for sustainable curated scientific resources. Sci. Data 11, 547. https://doi.org/10.1038/s41597-024-03406-w

Imteaz, M.A., Hossain, I., 2023. Climate change impacts on 'seasonality index' and its potential implications on rainwater savings. Water Resour. Manag. 37, 2593–2606. https://doi.org/10.1007/s11269-022-03320-z

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., Zhu, Z., 2017. Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. Glob. Change Biol. 23, 4133–4146. https://doi.org/10.1111/gcb.13787

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination. Agric. For. Meteorol. 121, 19–35. https://doi.org/10.1016/j.agrformet.2003.08.027

Kedron, P., Frazier, A.E., 2022. How to improve the reproducibility, replicability, and extensibility of remote sensing research. Remote Sens. 14, 5471. https://doi.org/10.3390/rs14215471

Kreshuk, A., Zhang, C., 2019. Machine learning: Advanced image segmentation using ilastik, in: Rebollo, E., Bosch, M. (Eds.), Computer Optimized Microscopy: Methods and Protocols. Springer, New York, NY, pp. 449–463. https://doi.org/10.1007/978-1-4939-9686-5_21

Kucharik, C.J., Norman, J.M., Gower, S.T., 1998. Measurements of branch area and adjusting leaf area index indirect measurements. Agric. For. Meteorol. 91, 69–88. https://doi.org/10.1016/S0168-1923(98)00064-1

Li, H., Zhou, Y., Zhao, X., Zhang, X., Liang, S., 2024. A dataset of 0.05-degree leaf area index in China during 1983–2100 based on deep learning network. Sci. Data 11, 1122. https://doi.org/10.1038/s41597-024-03948-z

Li, X., Du, H., Mao, F., Zhou, G., Han, N., Xu, X., Liu, Y., Zhu, D., Zheng, J., Dong, L., Zhang, M., 2019. Assimilating spatiotemporal MODIS LAI data with a particle filter algorithm for improving carbon cycle simulations for bamboo forest ecosystems. Sci. Total Environ. 694, 133803. https://doi.org/10.1016/j.scitotenv.2019.133803

Luiselli, L., Fa, J.E., 2019. Ecology and conservation of west african forests: An introduction. Acta Oecologica, Guinean Forests of West and Central Africa: Ecology and Conservation 94, 1–2. https://doi.org/10.1016/j.actao.2018.04.004

Ma, Y., Wang, W., Jin, S., Li, Haoxin, Liu, B., Gong, W., Fan, R., Li, Hui, 2023. Spatiotemporal variation of LAI in different vegetation types and its response to climate change in China from 2001 to 2020. Ecol. Indic. 156, 111101. https://doi.org/10.1016/j.ecolind.2023.111101

Malhado, A.C.M., Costa, M.H., de Lima, F.Z., Portilho, K.C., Figueiredo, D.N., 2009. Seasonal leaf dynamics in an amazonian tropical forest. For. Ecol. Manag. 258, 1161–1165. https://doi.org/10.1016/j.foreco.2009.06.002

Malhi, Y., Girardin, C., Metcalfe, D.B., Doughty, C.E., Aragão, L.E.O.C., Rifai, S.W., Oliveras, I., Shenkin, A., Aguirre-Gutiérrez, J., Dahlsjö, C.A.L., Riutta, T., Berenguer, E., Moore, S., Huasco, W.H., Salinas, N., Da Costa, A.C.L., Bentley, L.P., Adu-Bredu, S., Marthews, T.R., Meir, P., Phillips, O.L., 2021. The global ecosystems monitoring network: Monitoring ecosystem productivity and carbon cycling across the tropics. Biol. Conserv. 253, 108889. https://doi.org/10.1016/j.biocon.2020.108889

Nair, A., Ajith Joseph, K., Nair, K.S., 2014. Spatio-temporal analysis of rainfall trends over a maritime state (kerala) of India during the last 100 years. Atmos. Environ. 88, 123–132. https://doi.org/10.1016/j.atmosenv.2014.01.061

Obels, P., 2021. Monitoring plant growth in vertical farms using computer vision. Eindhoven University of Technology.

Prance, G.T., 1984. The vegetation of africa. by F. White. Brittonia 36, 273–273. https://doi.org/10.2307/2806524

Pyare, P.P., Henning, B., 2012. An efficient approach to standardizing the processing of hemispherical images for the estimation of forest structural attributes. Agric. For. Meteorol. 160, 1–13. https://doi.org/10.1016/j.agrformet.2012.02.007

Sarkar, D., Chapman, C.A., 2021. The smart forest conundrum: Contextualizing pitfalls of sensors and AI in conservation science for tropical forests. Trop. Conserv. Sci. 14, 19400829211014740. https://doi.org/10.1177/19400829211014740

Sumida, A., Watanabe, T., Miyaura, T., 2018. Interannual variability of leaf area index of an evergreen conifer stand was affected by carry-over effects from recent climate conditions. Sci. Rep. 8, 13590. https://doi.org/10.1038/s41598-018-31672-3

Veryard, R., Wu, J., O'Brien, M.J., Anthony, R., Both, S., Burslem, D.F.R.P., Chen, B., Fernandez-Miranda Cagigal, E., Godfray, H.C.J., Godoong, E., Liang, S., Saner, P., Schmid, B., Sau Wai, Y., Xie, J., Reynolds, G., Hector, A., 2023. Positive effects of tree diversity on tropical forest restoration in a field-scale experiment. Sci. Adv. 9, eadf0938. https://doi.org/10.1126/sciadv.adf0938

Wang, Y.-P., Zhang, L., Liang, X., Yuan, W., 2024. Coupled models of water and carbon cycles from leaf to global: A retrospective and a prospective. Agric. For. Meteorol. 358, 110229. https://doi.org/10.1016/j.agrformet.2024.110229

Wu, J., Albert, L.P., Lopes, A.P., Restrepo-Coupe, N., Hayek, M., Wiedemann, K.T., Guan, K., Stark, S.C., Christoffersen, B., Prohaska, N., Tavares, J.V., Marostica, S., Kobayashi, H., Ferreira, M.L., Campos, K.S., Da Silva, R., Brando, P.M., Dye, D.G., Huxman, T.E., Huete, A.R., Nelson, B.W., Saleska, S.R., 2016. Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. Science 351, 972–976. https://doi.org/10.1126/science.aad5068

Wu, J., Serbin, S.P., Xu, X., Albert, L.P., Chen, M., Meng, R., Saleska, S.R., Rogers, A., 2017. The phenology of leaf quality and its within-canopy variation is essential for accurate modeling of photosynthesis in tropical evergreen forests. Glob. Change Biol. 23, 4814–4827. https://doi.org/10.1111/gcb.13725

Wu, X., Niu, C., Liu, X., Hu, T., Feng, Y., Zhao, Y., Liu, S., Liu, Z., Dai, G., Zhang, Y., Van Meerbeek, K., Wu, J., Liu, L., Guo, Q., Su, Y., 2024. Canopy structure regulates autumn phenology by mediating the microclimate in temperate forests. Nat. Clim. Change 1–7. https://doi.org/10.1038/s41558-024-02164-2

Xie, X., Yang, Y., Li, W., Liao, N., Pan, W., Su, H., 2023. Estimation of leaf area index in a typical northern tropical secondary monsoon rainforest by different indirect methods. Remote Sens. 15, 1621. https://doi.org/10.3390/rs15061621

Y. Tong, Z. Jiao, X. Zhang, S. Yin, J. Guo, 2022. An improved method for estimating clumping index by digital hemispheric photography with field measurements. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2022.3216274

Yamaguchi, T., Tanaka, Y., Imachi, Y., Yamashita, M., Katsura, K., 2020. Feasibility of combining deep learning and RGB images obtained by unmanned aerial vehicle for leaf area index estimation in rice. Remote Sens. 13, 84. https://doi.org/10.3390/rs13010084

Yan, G., Hu, R., Luo, J., Weiss, M., Jiang, H., Mu, X., Xie, D., Zhang, W., 2019. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. Agric. For. Meteorol. 265, 390–411. https://doi.org/10.1016/j.agrformet.2018.11.033

Yan, K., Wang, J., Peng, R., Yang, K., Chen, X., Yin, G., Dong, J., Weiss, M., Pu, J., Myneni, R.B., 2024. HiQ-LAI: A high-quality reprocessed MODIS leaf area index dataset with better spatiotemporal consistency from 2000 to 2022. Earth Syst. Sci. Data 16, 1601–1622. https://doi.org/10.5194/essd-16-1601-2024

Yan, K., Zou, D., Yan, G., Fang, H., Weiss, M., Rautiainen, M., Knyazikhin, Y., Myneni, R.B., 2021. A bibliometric visualization review of the MODIS LAI/FPAR products from 1995 to 2020. J. Remote Sens. 2021. https://doi.org/10.34133/2021/7410921

Yang, S., Peng, N., Zhai, D., Tao, Y., He, Q., Mu, X., Li, Y., Fan, W., 2023. Fisheye-Based Forest LAI Field Measurements for Remote Sensing Validation at High Spatial Resolution. IEEE Geosci. Remote Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3308369

Zhang-Zheng, H., Adu-Bredu, S., Duah-Gyamfi, A., Moore, S., Addo-Danso, S.D., Amissah, L., Valentini, R., Djagbletey, G., Anim-Adjei, K., Quansah, J., Sarpong, B., Owusu-Afriyie, K., Gvozdevaite, A., Tang, M., Ruiz-Jaen, M.C., Ibrahim, F., Girardin, C.A.J., Rifai, S., Dahlsjö, C.A.L., Riutta, T., Deng, X., Sun, Y., Prentice, I.C., Oliveras Menor, I., Malhi, Y., 2024a. Contrasting carbon cycle along tropical forest aridity gradients in west africa and amazonia. Nat. Commun. 15, 3158. https://doi.org/10.1038/s41467-024-47202-x

Zhang-Zheng, H., Deng, X., Aguirre-Gutiérrez, J., Stocker, B.D., Thomson, E., Ding, R., Adu-Bredu, S., Duah-Gyamfi, A., Gvozdevaite, A., Moore, S., Oliveras Menor, I., Prentice, I.C., Malhi, Y., 2024b. Why models underestimate west african tropical forest primary productivity. Nat. Commun. 15, 9574. https://doi.org/10.1038/s41467-024-53949-0

Zhu, L., Guan, P., Liu, W., Zheng, Y., 2023. Optimization of lai estimation method based on smartphones with fisheye lens. Int. J. Circuits Syst. Signal Process. 17, 112–122. https://doi.org/10.46300/9106.2023.17.14

Zou, L., Stan, K., Cao, S., Zhu, Z., 2023. Dynamic global vegetation models may not capture the dynamics of the leaf area index in the tropical rainforests: A data-model intercomparison. Agric. For. Meteorol. 339, 109562. https://doi.org/10.1016/j.agrformet.2023.109562

Acknowledgement

We thank Vivian Fe Ursula Schader, Ina Pelster, and Katrin Wilhelm for their trail of the protocol and insightful comments. M.T. and L.F. was supported by Second-class Scholarship (by China Agricultural University, Beijing). H.Z-Z was supported as part of the Next Generation Ecosystem Experiments-Tropics, funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research. This work is a product of the Global Ecosystems Monitoring (GEM) network (gem.tropicalforests.ox.ac.uk). Fieldwork was funded by grants from the UK Natural Environment Research Council (NE/I014705/1 and NE/P001092/1) to Y.M. and Africa Oxford Initiative Catalyst Grant to H.Z. We also

acknowledge the Wildlife Division of the Forestry Commission in Ghana as well as the many field assistants who helped with data collection from the field.

Author contributions

M.T., H.Z., conceived the study. S.A.B., A.D.G., H.Z., Y.M. collected field data. M.T., L.F.N., T.M.X, H.Z., did the modelling work. M.T., T.M.X, H.Z., carried out data analysis. H.Z., C.D.,Y.M., revised the manuscript. M.T. wrote the paper with inputs and revisions from all co-authors.

Competing interest statement

The authors declare no competing interests.

Inclusion & Ethics statement

Local researchers were involved throughout the research process, with clear roles and responsibilities agreed upon, particularly with contributions from the Forestry Research Institute of Ghana. The research was locally relevant and determined in collaboration with local partners. Capacity-building workshops were organized during the research process. Local and regional research relevant to the study is cited, as evidenced by the references to West Africa specific studies. Risk management plans were made for the fieldwork following the universities' regulations.

Leaf area index protocol, introduced by 'An AI-based and coding-free protocol for forests leaf area index calculation'

Tao Ma1,2, Minxue Tang3, Fangni Liu1, Akwasi Duah-Gyamfi4, Stephen Adu-Bredu4, Mark Arcebal K. Naive5,6, Imma Oliveras Menor2,7, Sam Moore2, Zhiyuan Zhang8, Yadvinder Malhi2,9✉, Cecilia A. L. Dahlsjö2,9, Huanyuan Zhang-Zheng2,9✉
(yadvinder.malhi@ouce.ox.ac.uk) (huanyuan.zhang@ouce.ox.ac.uk)

1 College of Resources and Environmental Sciences, China Agricultural University, Beijing, China

2 Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom

3 Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, United Kingdom

4 Forestry Research Institute of Ghana, Council for Scientific and Industrial Research, Kumasi, Ghana

5 Center for Integrative Conservation and Yunnan Key Laboratory for Conservation of Tropical Rainforests & Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China

6 University of Chinese Academy of Sciences, Beijing 100049, China

7 AMAP – botAnique et Modélisation de l'Architecture des Plantes et des Végétations, Université de Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier CEDEX 5, France

8 College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China

9 Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, United Kingdom

Taking Hemispherical photo in the field

The accuracy of exposure settings is crucial when using digital hemispherical photography (DHP) to measure canopy structure. Chen et al. (1991) was among the first to highlight that automatic exposure can introduce bias by reducing contrast between the sky and canopy. Later, Zhang et al. (2005) recommended adjusting manual settings—such as increasing exposure by 1–2 stops compared to open-sky conditions—to enhance measurement accuracy. Numerous subsequent studies have explored methods for camera exposure parameterization (highlighted in red). However, a recent study acknowledged that in practice—particularly in tropical forests—manual benchmarking against open-sky conditions is often impractical due to time and accessibility constraints (Jiao et al., 2023). The study proposed an innovative solution: capturing multiple images at varying exposures and merging them using image fusion techniques (e.g., Sony's auto HDR function if you own Sony camera) to optimize canopy detail. While the researchers demonstrated the effectiveness of this approach, implementation remains challenging: no user-friendly software exists yet, and the requirement to capture multiple shots per location is time-intensive.

Here's our practical advice: If you have a strong grasp of those papers and sufficient time, we recommend setting manual exposure or experimenting with multi-exposure fusion. However, if these methods are impractical in your field setting, don't be overly concerned—just avoid extreme cases. For instance ensure the sky isn't overly bright (Figure 1), as this can mask leaf details. For modern cameras (especially Sony models), enable Auto HDR and Auto Exposure, but familiarize yourself with brightness adjustments in your camera settings for necessary finetuning. Aim for good exposure, but avoid letting perfectionism hinder fieldwork progress. This is especially true if you're using our AI-based workflow, as minor exposure variations have minimal impact on LAI results. Ideal lighting conditions include diffuse light, such as overcast skies or dawn/dusk periods (Chen et al., 1991). That said, in tropical forests—where lighting can be exceptionally dim—the environment may become too dark for clear shots. In such cases, prioritize shutter speed and focus—image clarity is far more critical than exposure. Discard any blurry images during post-processing (Figure 3), and never use a flash in low-light conditions.

Figure 1

It is essential to use a camera equipped with a circular fisheye lens that provides a full 180-degree field of view (so when pointed upright, the photo's edges capture the ground). Avoid diagonal fisheye lenses, as they typically offer only a 180-degree view along the diagonal, resulting in incomplete canopy coverage. If purchasing a dedicated camera is not feasible, a smartphone such as an iPhone can serve as an alternative. However, you must attach a 180-degree circular fisheye lens to your device and verify that it is truly circular (not diagonal).

For fieldwork in tropical forest plots measuring 1 hectare, the standard practice is to capture hemispherical photographs at 25 evenly distributed points each month (Xie et al., 2023). After collecting the initial dataset, you can experiment by analyzing subsets of these images to determine whether a reduced number of photographs still produces consistent LAI estimates, which may allow for fewer photos in future surveys. For camera orientation, mount the camera on a tripod and ensure it is perfectly level. The lens should point directly upward, and aligning the camera to magnetic north helps ensure consistent data analysis. Permanently mark each photo location point so the same points can be reliably relocated during subsequent surveys. In tropical forests, additional equipment such as a tree lopper and sickle may be required (Figure 2).

Figure 2 Remember to bring a tree lopper, we don't want big leaves that obscure most of the image.

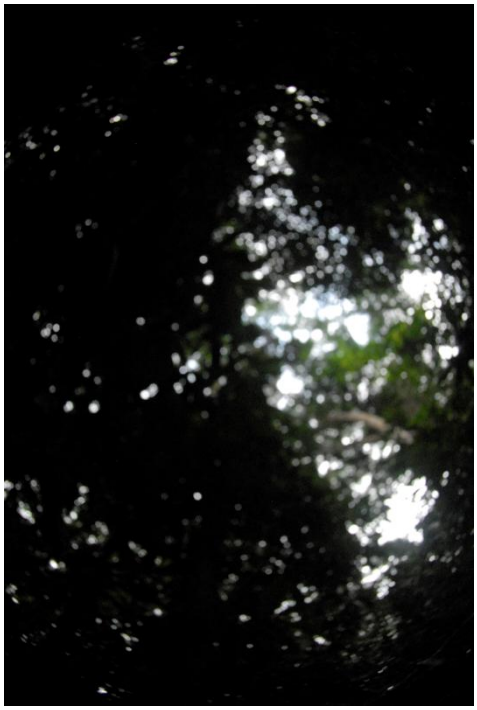


Figure 3 Plot ANK-01, taken in July 2013. Using manual focus and a large aperture may help to get clear focus, but still in a dark tropical forest some photos may lose focus if not carefully checked in the field.

Chen, J.M., Black, T.A., Adams, R.S., 1991. Evaluation of hemispherical photography for determining plant area index and geometry of a forest stand. Agric. For. Meteorol. 56, 129–143. https://doi.org/10.1016/0168-1923(91)90108-3

Jiao, S., Chen, Y., Cheng, Y., Duan, T., Gao, Z., Sun, Y., Huang, F., 2023. MEF-DHP: Digital Hemispheric Photography Method Based On Multi-Exposure Fusion, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. Presented at the IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, pp. 2755–2758. https://doi.org/10.1109/IGARSS52108.2023.10282038

Xie, X., Yang, Y., Li, W., Liao, N., Pan, W., Su, H., 2023. Estimation of Leaf Area Index in a Typical Northern Tropical Secondary Monsoon Rainforest by Different Indirect Methods. Remote Sens. 15, 1621. https://doi.org/10.3390/rs15061621

Zhang, Y., Chen, J.M., Miller, J.R., 2005. Determining digital hemispherical photograph exposure for leaf area index estimation. Agric. For. Meteorol. 133, 166–181.

Using ilastik to process raw fisheye images

Note that you must right-click ‘ilastik’ and select ‘Run as administrator’ on Windows computer. Ilastik may take up to 5 minutes to initialize. On Windows, you might need to press Enter when the first black screen.

First, create a new project using the Pixel Classification module (Figure 4).

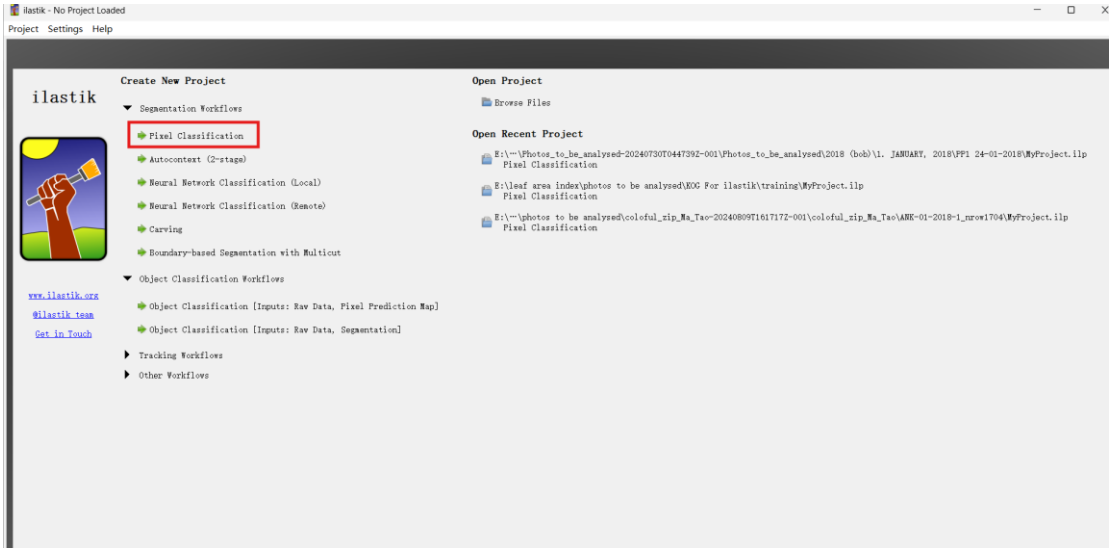


Figure 4

Next, choose a folder to store the project. Ensure the project is saved in the same folder as the training data before clicking save (Figure 5).

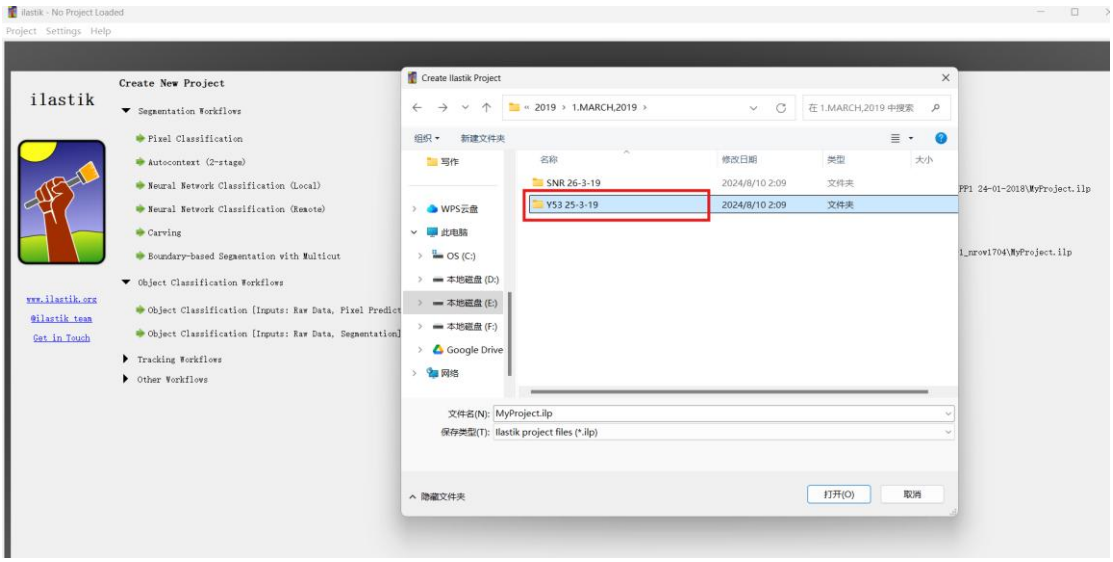


Figure 5

In the Input Data section, select images for analysis by clicking Add New, then click Add Separate Images (Figure 6). For example, the folder “Y53 25-3-19” would contain your training data (in this case, 25 images). I suggest naming it ‘training_data’ instead of Y53. It’s recommended to select contrasting photos from your photo pool and copy them into this folder. Typically, I include photos from various sites, seasons and weather conditions to ensure they represent the full range of data. A good starting point is 10 photos. Note: If you plan to share your ilastik model, you must send the entire “Y53 25-3-19” folder, not just the .ilp file. After saving, you’ll notice a new .lip file in this folder (this file stores your manual annotations, as explained later).

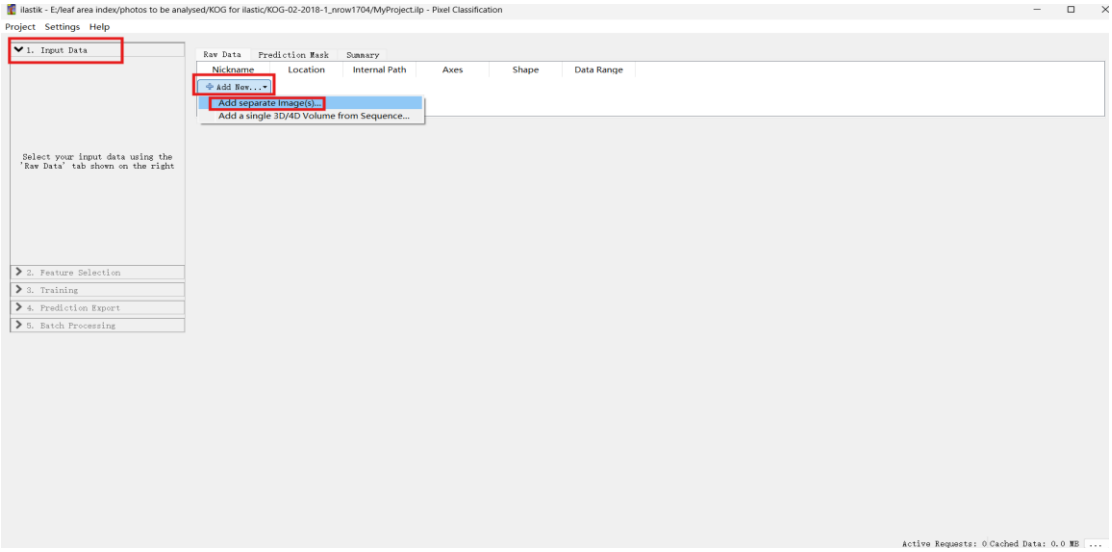


Figure 6

Next, go to the folder that contains all the training data. You can either drag the files or press Ctrl+A to select all photos at once (Figure 7).

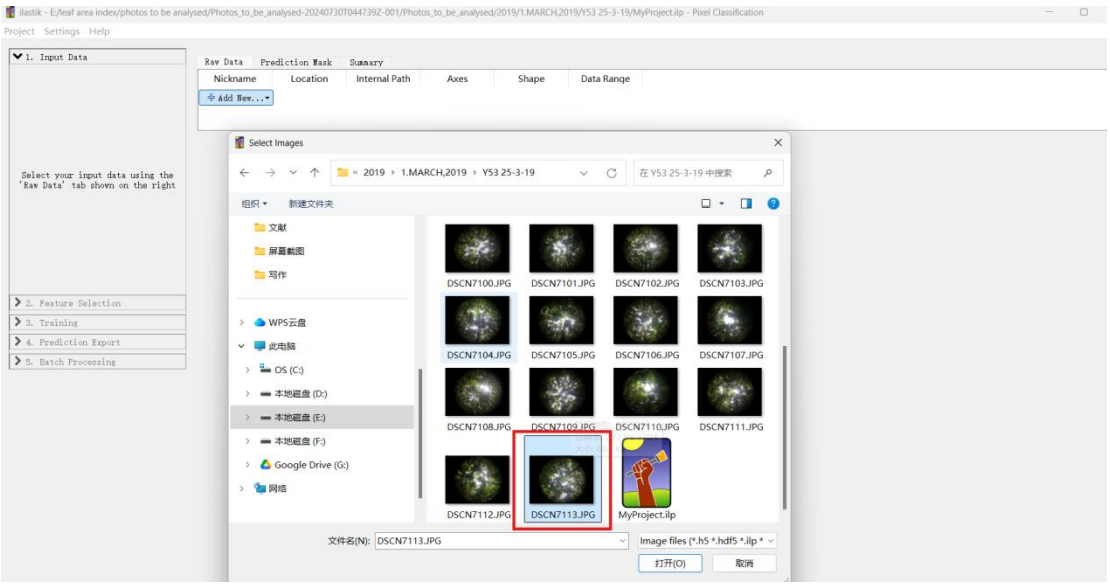


Figure 7

Alternatively, you may add only one photo now and come back later to add more if additional training photos are needed (Figure 8 and Figure 9). Naturally, you can delete those any photos that aren't used for training (after completing the protocol).

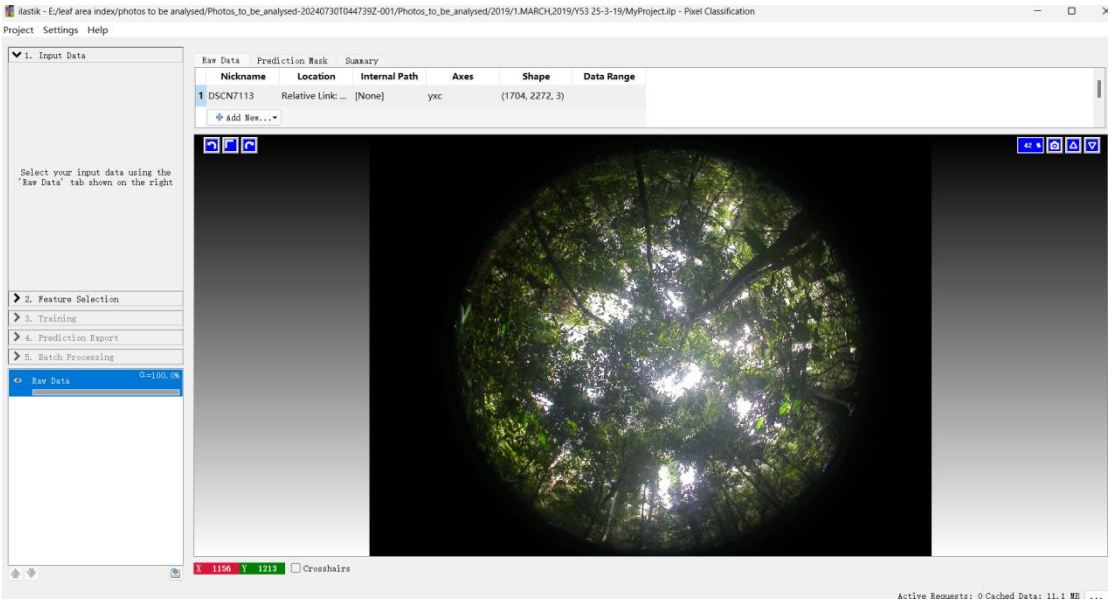


Figure 8

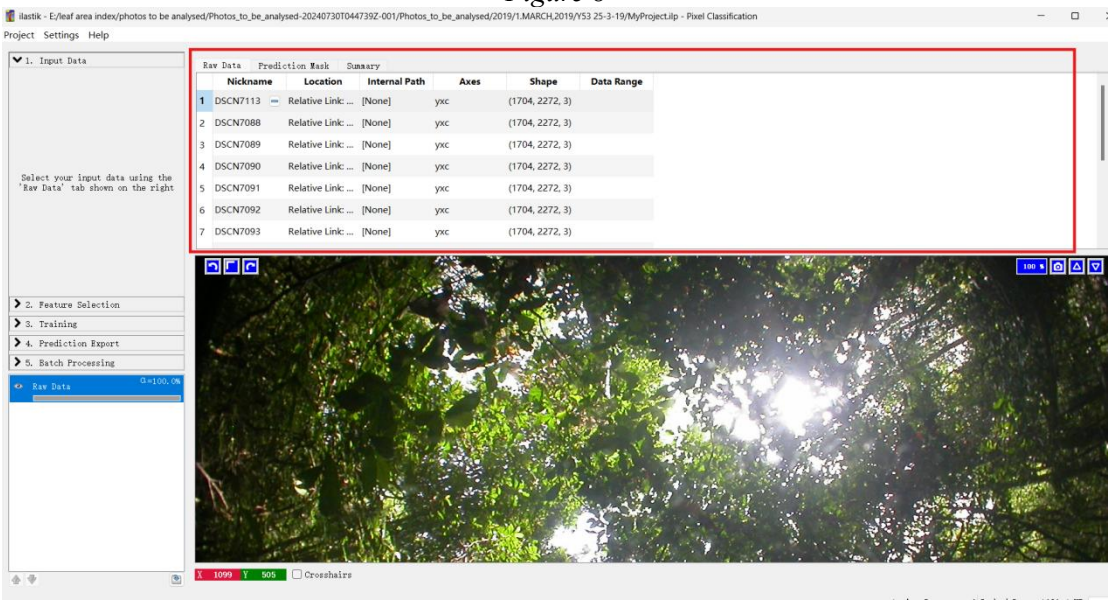


Figure 9

Next, we will go to feature selection. Click “Select Features” to proceed (Figure 10).

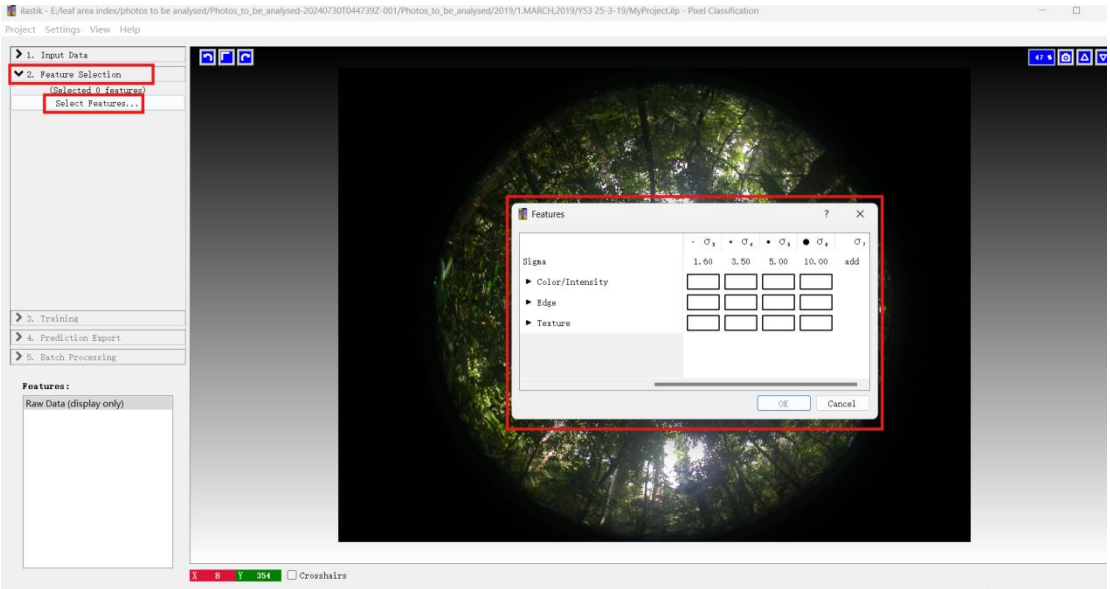


Figure 10

The ‘Features’ window in ilastik is dedicated to feature selection (Figure 11). These parameters allow users to configure various feature filters, enabling the classification algorithm to more effectively recognize and process image characteristics such as textures, edges, colors, and other relevant information.

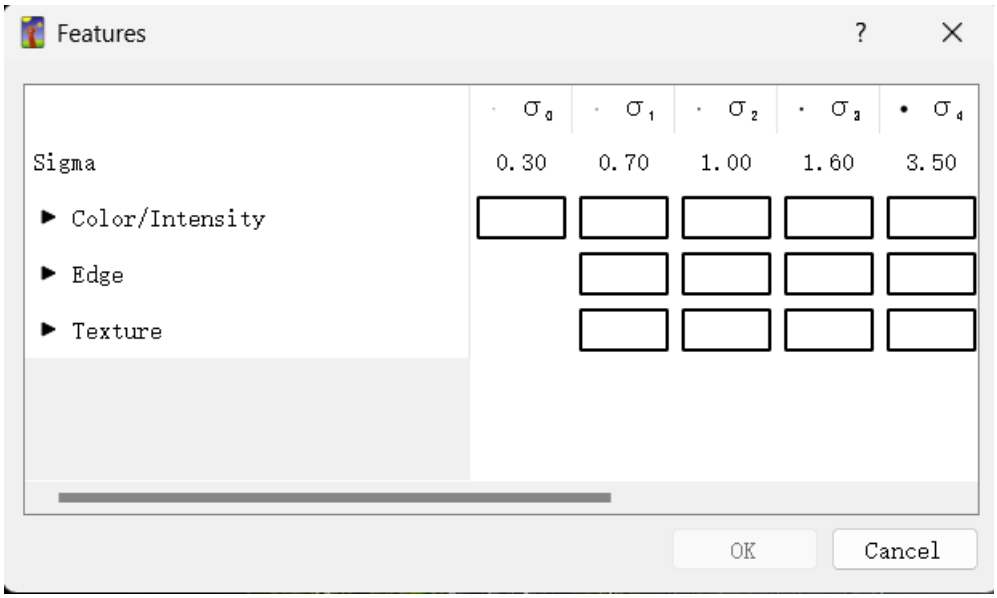


Figure 11

The following functions are available:

(1) Color/Intensity: Color is one of the most distinguishing features for differing between the canopy and the sky. Because of leaves, branches, and shadows, the canopy typically appears in shades of green, brown, or darker tones, while during sunrise or sunset, the sky often appears in lighter colors, such as shades of blue, white, or even

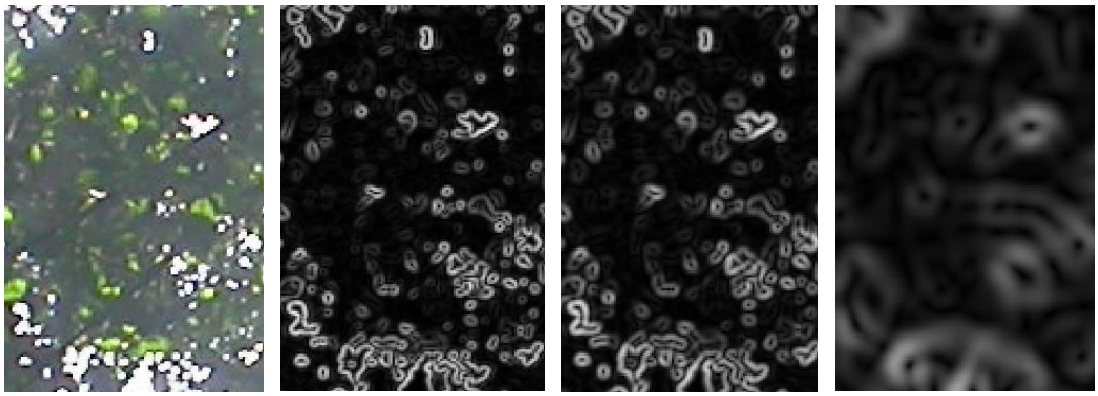
orange and pink.

(2) Edge: Edge features detect boundaries characterized by significant changes in pixel intensity or color.

- Canopy-Sky Boundary: The irregular outlines of leaves and branches against the sky create distinct edges.
- Complexity: The natural shapes of the canopy introduce complex edges that enhance recognized.

(3) Texture: Texture captures the surface patterns and variations within image regions. Owing to the arrangement of leaves, branches, and varying shadows, the canopy exhibits a complex texture, while the sky typically has a smooth texture, though clouds can introduce some patterns.

Each feature can be applied at various scales, with the scale corresponding to the sigma of the Gaussian function used to blur the image before applying the filter. A higher sigma value allows the filter to capture information from a wider area while simultaneously reducing finer details. The following image provides an example of an edge filter computed using three different sigma values. For example, filters with lower sigma values detect small edges more effectively, while those with higher sigma values identify only more general or broader structures.



Raw data | Sigma=0.7 | Sigma=1 | Sigma=1

To enable specific image features (e.g., color, edge, texture) and assign their corresponding Sigma values, activate the respective checkbox (green check mark). The selected features are incorporated into the machine learning model and utilized to enhance class discrimination during both the training and classification phases. For this analysis, I will enable the following features (Figure 12). While additional selections are possible, they may increase computational processing time.

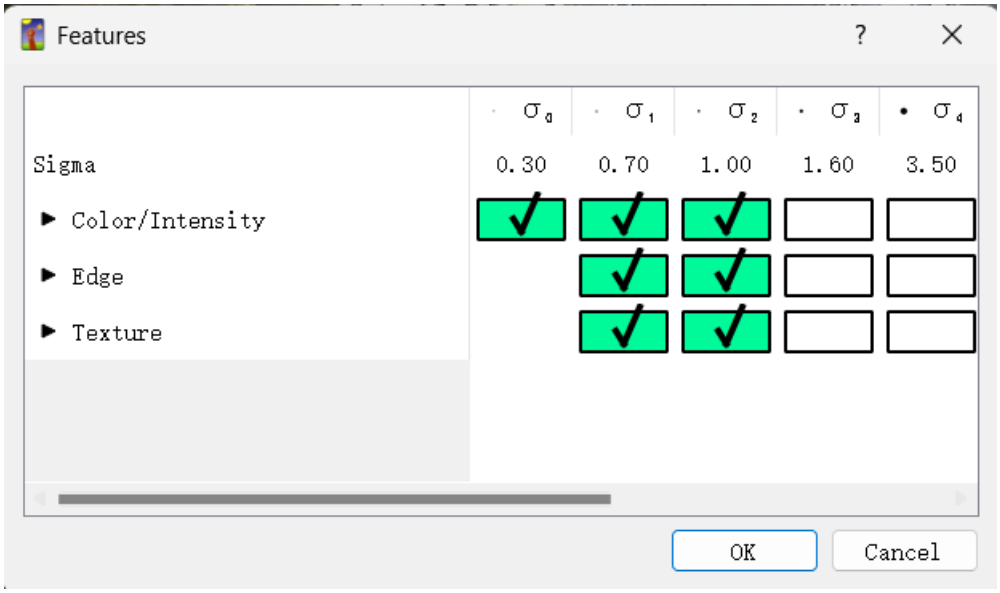


Figure 12

The complete set of selected features is displayed in this panel (Figure 13). Here, users can interact with individual features by clicking on them to assess their discriminatory power—specially, how effectively each feature differentiates the target objects (leaves) from background elements (sky). This visualization step enables qualitative validation of feature selection prior to model training.

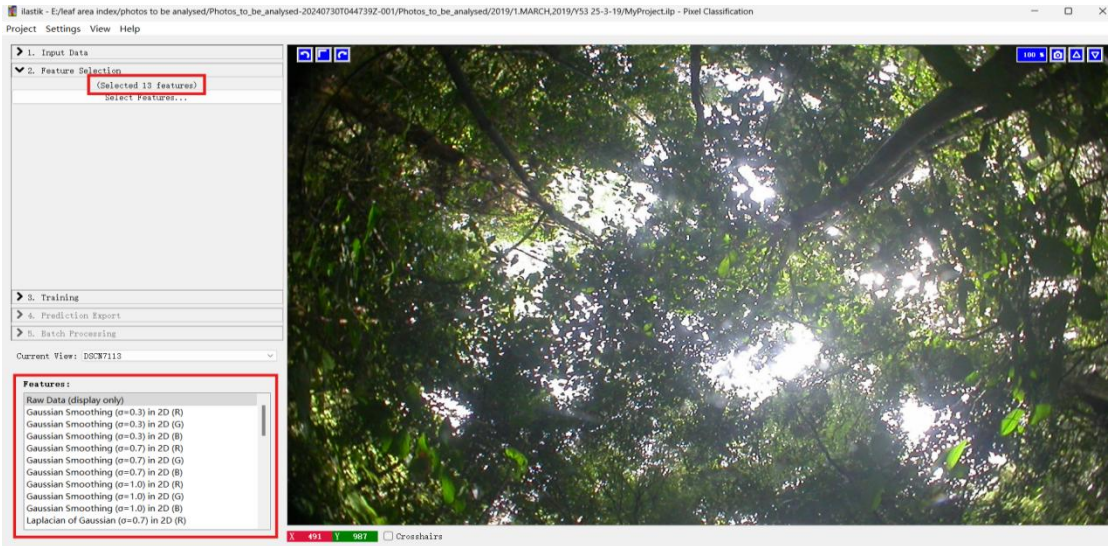


Figure 13

The subsequent phase of pixel classification entails training discriminative model to differentiate between distinct object categories. This iterative workflow requires the user to: (1) provide initial annotations for representative regions, (2) evaluate the model's predictive output, and (3) iteratively refine the annotations to correct misclassified pixels and improve the model accuracy. Each iteration enhances thew model's ability to generalize from the annotated examples.

We now proceed to the training phase within ilastik, which requires manual annotation of image regions to differentiate between sky and leaf categories (Figure 14). Label customization is achieved double-clicking the color swatch adjacent to each class designation where the default configuration assigns blue (first label) to sky regions and green (second label) to leaves. Critical implementation note: The ordinal position of labels directly determines the output values in the processed images, where the first category is encoded as 1st (sky) and 2nd (leaves). Additionally, enabling renormalization option [min, max] during subsequent processing stages may introduce alternate pixel values (0 or 100) in the exported results.

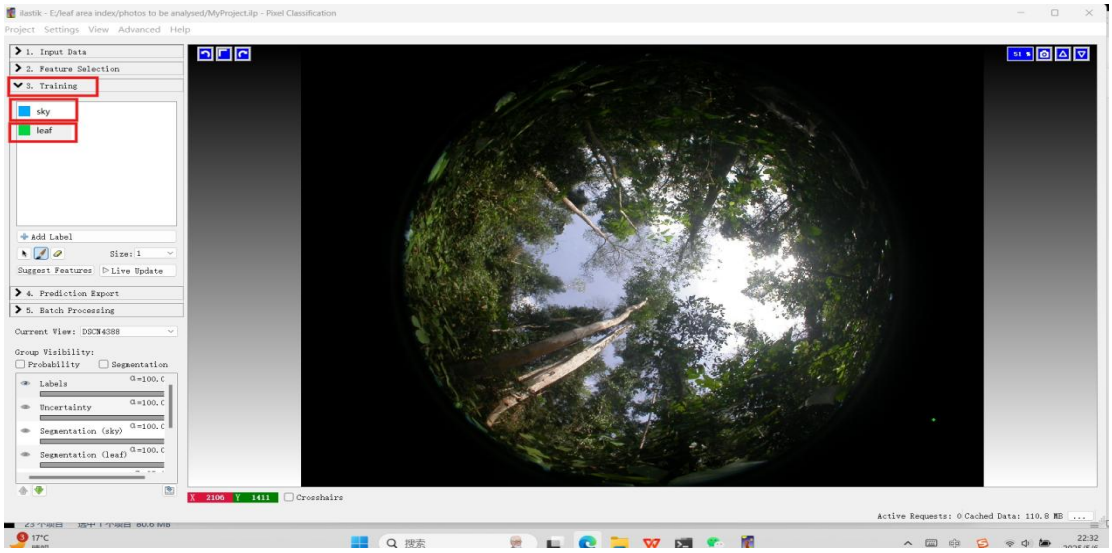


Figure 14

Here, we use two labels—blue for sky and green for leaves (Figure 15). Note that CAN-EYE prefers sky as the first category, so we recommend assigning as Label 1 to ensure consistency with standard processing.

Figure 15

For particularly complex forest canopies, such as those with dense branching structures, a ternary classification approach may be applied. In this method, branches are labeled in red to explicitly exclude them from leaf classification (Figure 16).

Additionally, for "grey areas" where human judgement is uncertain (i.e., when distinguishing between branches and leaves is unclear), these areas should be left unmarked. The algorithm will automatically process these zones. Only confidently identifiable features should be annotated.

To adjust the view, hold the Ctrl key while scrolling to zoom, and use the middle mouse button to pan the image. Note that this function requires either a mouse or touchscreen device; touchpad controls are not supported.

Figure 16

Click Live Update to observe ilastik's classification results for your image (Figure 17). Toggle between probability view and segmentation view to examine the analysis output. If required, you may add manual annotations to improve results, though this should only be done when Live Update is disabled to prevent system instability. <u>Important Note: Always ensure Live Update is turned off before annotating, as concurrent operations may cause software crashes, and save your project regularly to prevent data loss.</u>

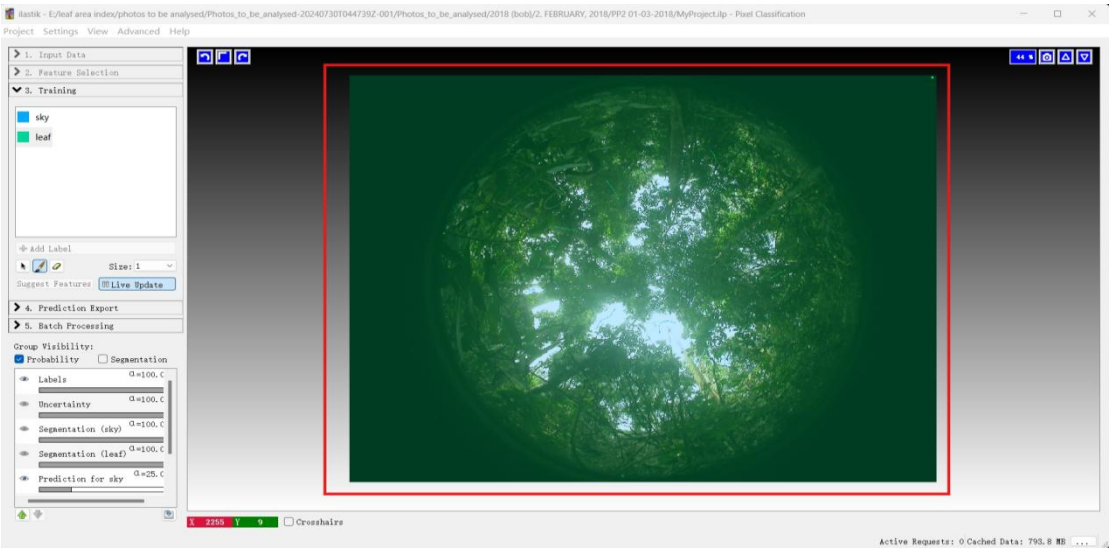


Figure 17

To view the final classification results, enable the Segmentation overlay by activating Segmentation checkbox in the interface (Figure 18).

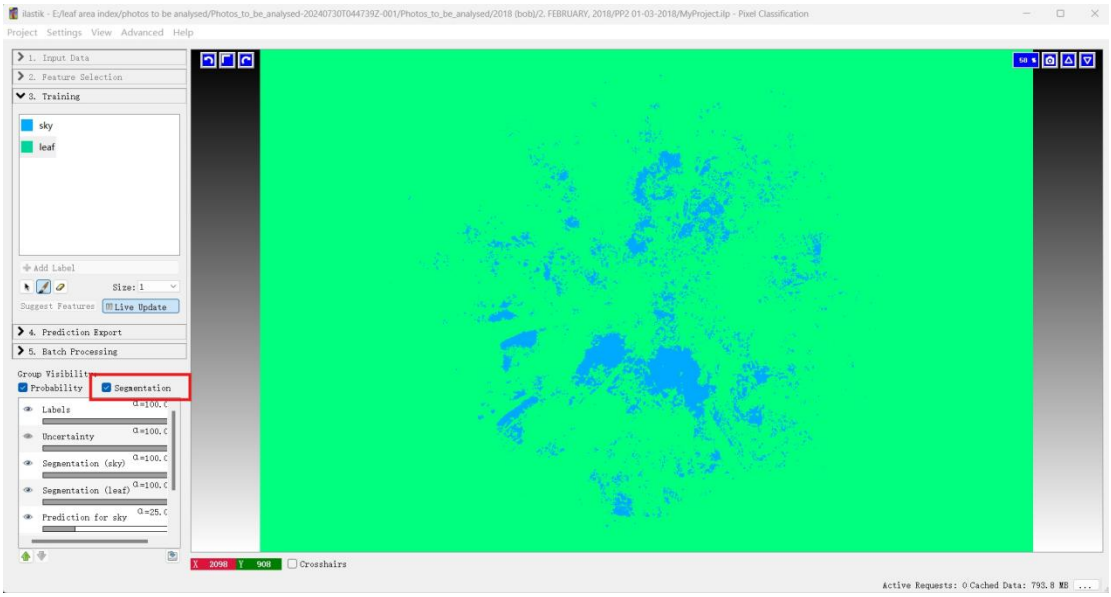


Figure 18

If the segmentation appears incomplete or inaccurate, pause Live Update and refine the annotations by adding more marks. Note that you can use Ctrl+Z (undo) or the eraser tool to correct any mistakes. Once you finish annotating the current image, click "Current View" to proceed to the next training image (additional images can be imported in the 1. Input Data step if needed). Re-enable Live Update—the software will now segment the new image using the knowledge gained from previous annotations. If the results are unsatisfactory, add more marks to improve segmentation. If the output meets expectations, move on to the next image. Repeat this iterative process until the model consistently produces satisfactory results without requiring additional manual

corrections on new images.

If the current results meet your quality standards , disable Live Update by clicking it again, then select Suggest Features to open a new configuration dialog box (Figure 19).

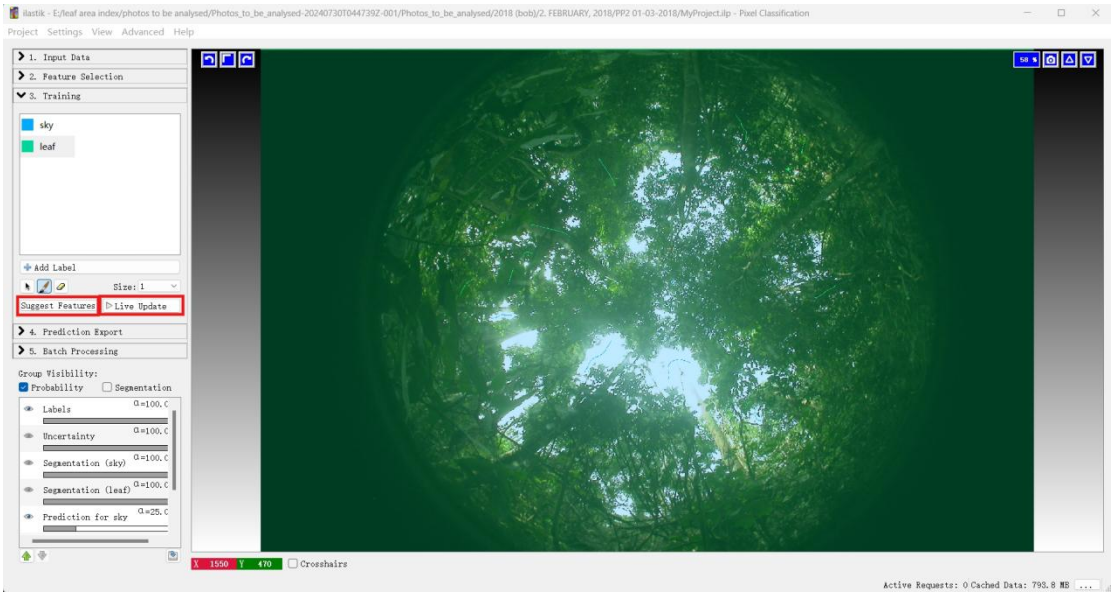


Figure 19

If you are uncertain about your pixel feature selections, the Suggest Features functionality can assist in optimizing them. This feature evaluates the classifier's performance on annotated pixels by testing different feature combinations. The process employs machine learning technique, including random forest classification and "out-of-bag" error estimation, to assess prediction accuracy. Additionally, ilastik provides computation time metrics, allowing users to balance runtime efficiency with segmentation quality. For users without a computer science background, the system automatically refines selection to improve both prediction accuracy and processing speed, reducing the need for manual tuning.

Next, you will reach the "Feature Selection" page. For users with machine learning expertise, this section provides detailed explanations of key functionalities and adjustable parameters, allowing you customized the process based on your requirements. For those without a technical background, you may safely ignore these advanced options and proceed by selecting the "Filter Method (recommended)" option. "Run Feature Selection" to continue (Figure 20).

Note: You will need to run ilastik as Administrator on Windows if you get an error

Figure 20

Select "Filter Method (recommended)" and then click “Run Feature Selection”. Choose then appropriate feature selection and click “Select Feature Set” (Figure 21).

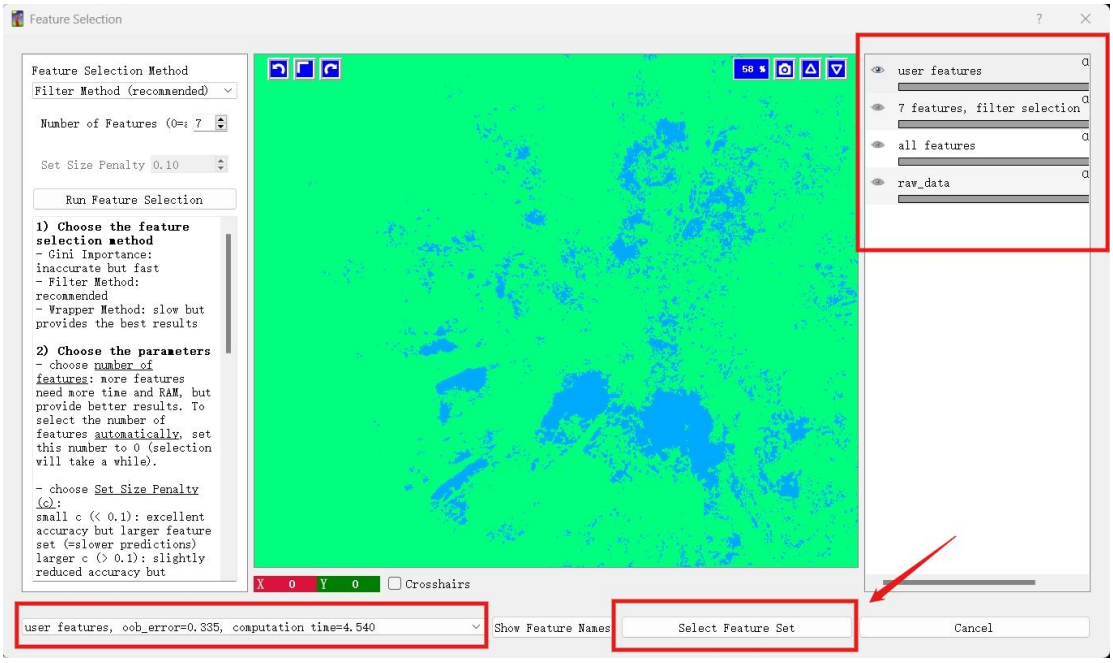


Figure 21

In the displayed ilastik feature selection interface, the three options highlighted in the red box at the bottom left are located in the drop-down menu below. Each option represents different feature selection methodologies and usage scenarios. Their key differences are as follows:

(1) User features: This denotes the feature set manually selected by the user during configuration.

(2) 7 features, Filter selection: This option indicates that 7 features were automatically selected using the Filter Method(recommended).

(3) all features: This option includes all available features without reduction (i.e., no filtering was applied), comprising the full set of initial features.

Notably, all three options provide two metrics:

- Oob Error (Out-of-bag Error): Quantifies the model's error rate when evaluated on unseen data, serving as a robust estimate of generalization performance.
- Computation Time: Represents the duration required for the model to process the corresponding feature set.

In an optimal scenario, both the oob_error and computation time should be minimize. However, practical implementation often requires trade-offs between these metrics. When computational resources are constrained or rapid results are needed, selecting fewer features may be optimal. Conversely, if model accuracy is the primary objective, increasing the feature count—despite longer computation times—may be justified.

For this demonstration, we select “User Features” as the most appropriate choice for our workflow. After selection, click “Select feature Set” to exit the Feature Selection interface, followed by “Live Update” to view the results. It is important to note that the “Suggest Feature” step can be optionally bypassed as it is particularly computationally demanding. This step may cause performance issues or system instability when processing extensive training datasets, numerous pre-selected features, or when running on hardware with limited resources such as standard laptops. In such cases, we recommend first reducing the number of pre-selected features, limiting the quantity of training images and annotations, then running the “Suggest Feature” process once to identify the optimal feature set for future use. If these adjustments still prove problematic, users can confidently proceed with the predefined recommended features and skip this step entirely, as it is not strictly necessary for successful implementation.

When we click “Select feature Set”, the Feature Selection screen closes. Then, click “Live Update” to view the results (Figure 22).

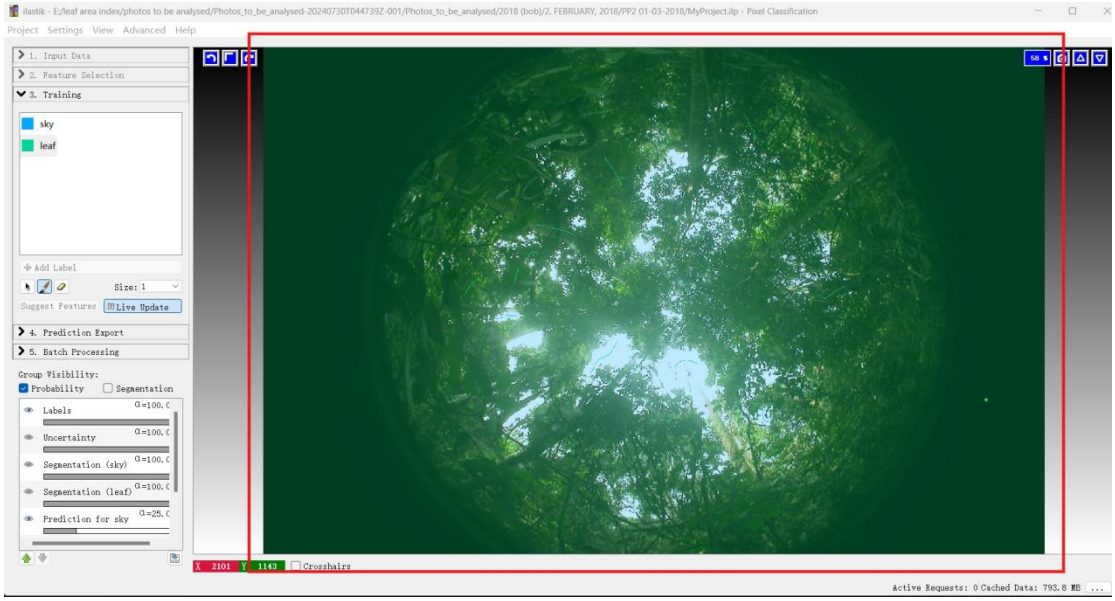


Figure 22

(NoteL: If you are using CAN-EYE, set “sky” as the 1st category.)

Stop the “Live Update” and navigate to 4, the “Prediction Export” section (Figure 23). Now that you're in “Prediction Exports”, locate the “Export Settings” panel and under “Source”, change the selection “Probabilities” to “Simple Segmentation”.

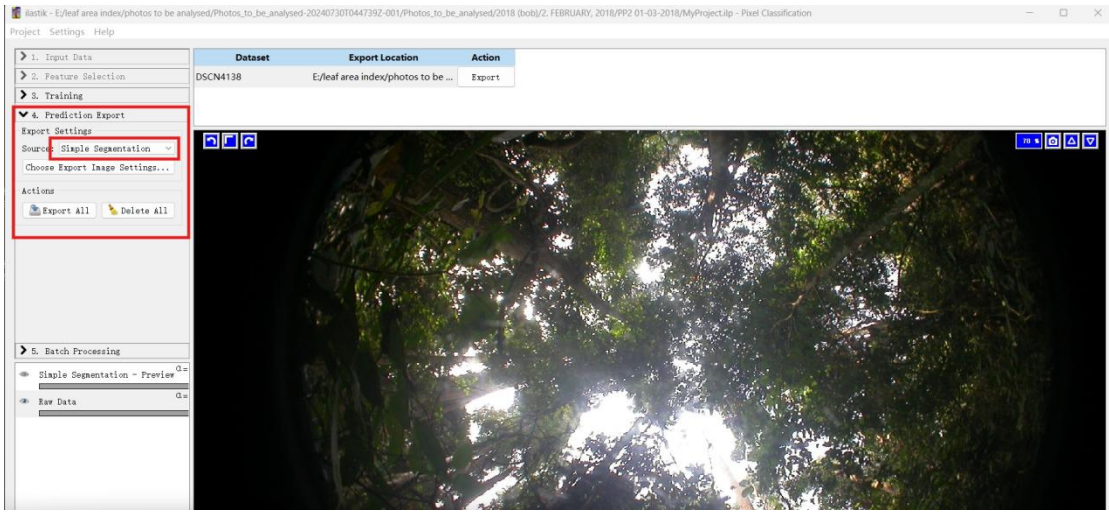


Figure 23

(Change Probabilities to Simple Segmentation)

First, Click ‘Choose Explore Image Settings' and under the 'Transformation', select 8-bit integer. Then enable ‘Renormalize’ and change the value from '1 to 2' to '0 to 100' (Figure 24). Note that in the training phase, label 1 represents the sky while label 2 represents the leaf. If you don’t enable Renormalize, ilastik will save output images

containing values 1 and 2. When you enable Renormalize, all pixels in your 1st category (label 1, which should be sky for CAN-EYE) will be converted to 0, while the 2nd category (label 2) will become 100. For ternary classification (when you have categories 1, 2, and 3), enabling Renormalize with '1 to 2' to '0 to 100' setting will convert category 1 to 0, category 2 to 100, and category 3 will become 100. In such cases, it's recommended to leave 'Renormalize' disabled, and instead use Python, R or Matlab to manually convert the output values (1, 2, 3) to your preferred numbers. To summarize the protocol for CAN-EYE: make sure your category 1 is assigned to sky, then enable Renormalize and set the conversion from '1 to 2' to '0 to 100'.

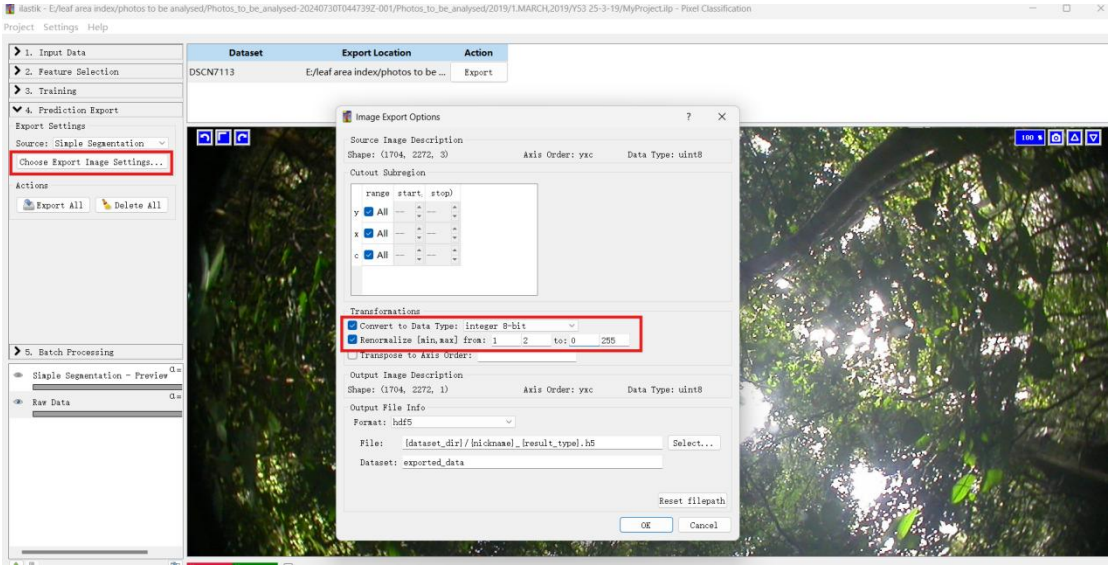


Figure 24

Under 'Output File Info', change the data type from hdf5 to jpeg, as this format easier to view. Then click OK (Figure 25).

Under files, the interface shows where your outputs will be saved. You can change this to another folder if needed, but remembering that the naming convention {nickname}_{result_type} is critical. For example, my output path is: C:/Outputs/{nickname}_{result_type}.jpg

Important: Always use forward slashes (////) in the path, not backslahes (\\\\).

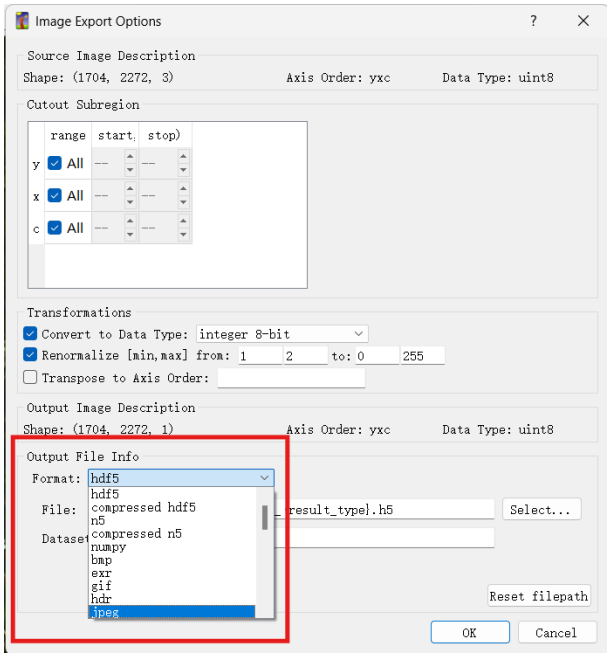


Figure 25

Then use 'Export' or 'Export All' to convert all your training photos to black and white (Figure 26).

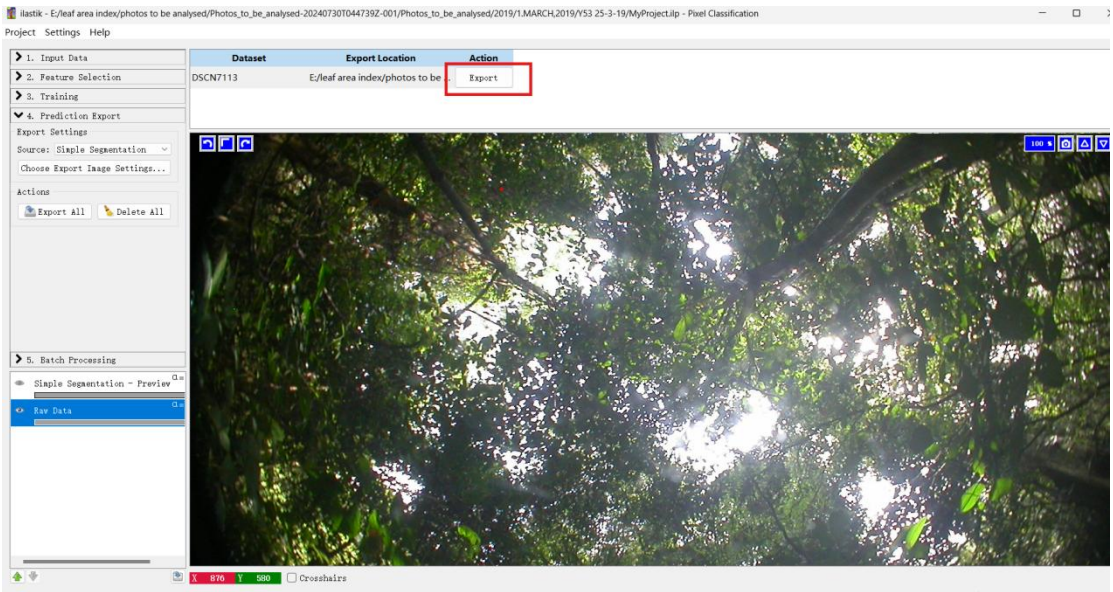


Figure 26

Don't forget to save your project regularly! Otherwise, you may encounter errors when you next open the ilastik file (Figure 27).

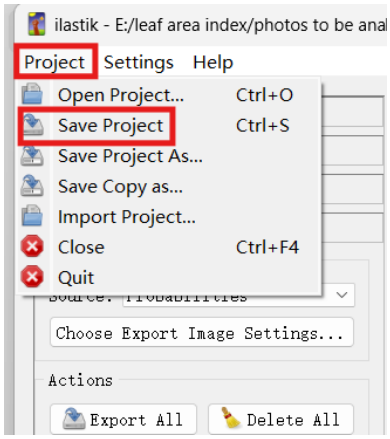


Figure 27

Now, you will find the black and grey image in your output folder (Figure 28).

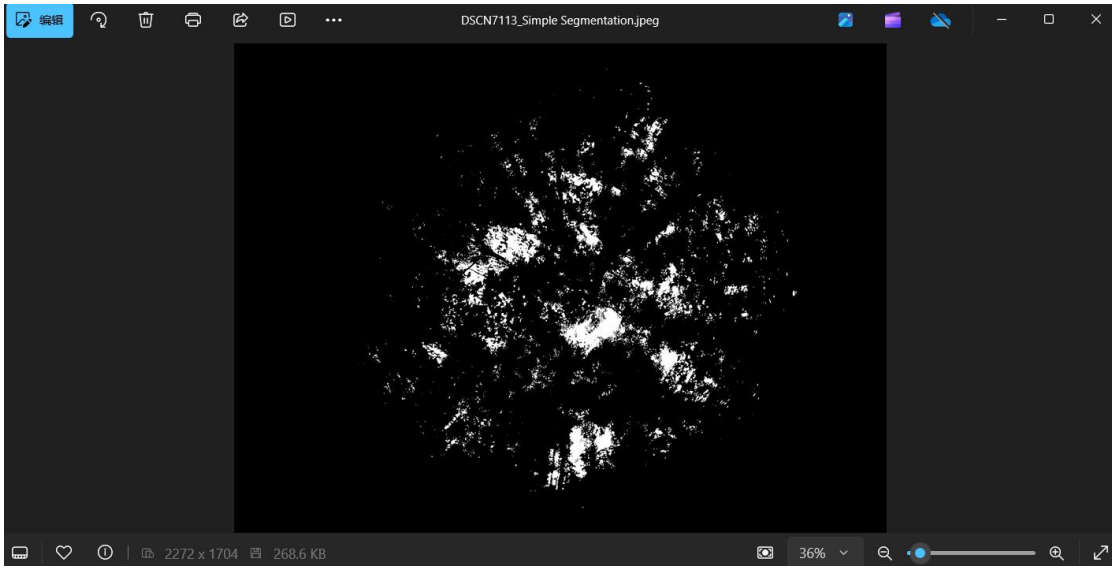


Figure 28

Because this is a JPEG, the values of 0 and 100 would be understood as greyscale – where 0 is black and 100 is grey. Thus, if you uncheck 'Renormalize', meaning you have values of 1 and 2, the image **will appear totally black**! To view such an image, I normally use QGIS software.

To get better results, you should include more images in your input data. Here I have 20 images, which is approximately the minimum number needed to get relatively good results when processing many images.

Now, you can use Batch Processing to process all your images (other than those used for training)(Figure 29).

Before batch processing, an **EXTREMELY** important step is to name your images

consistently so that you can identify them later.

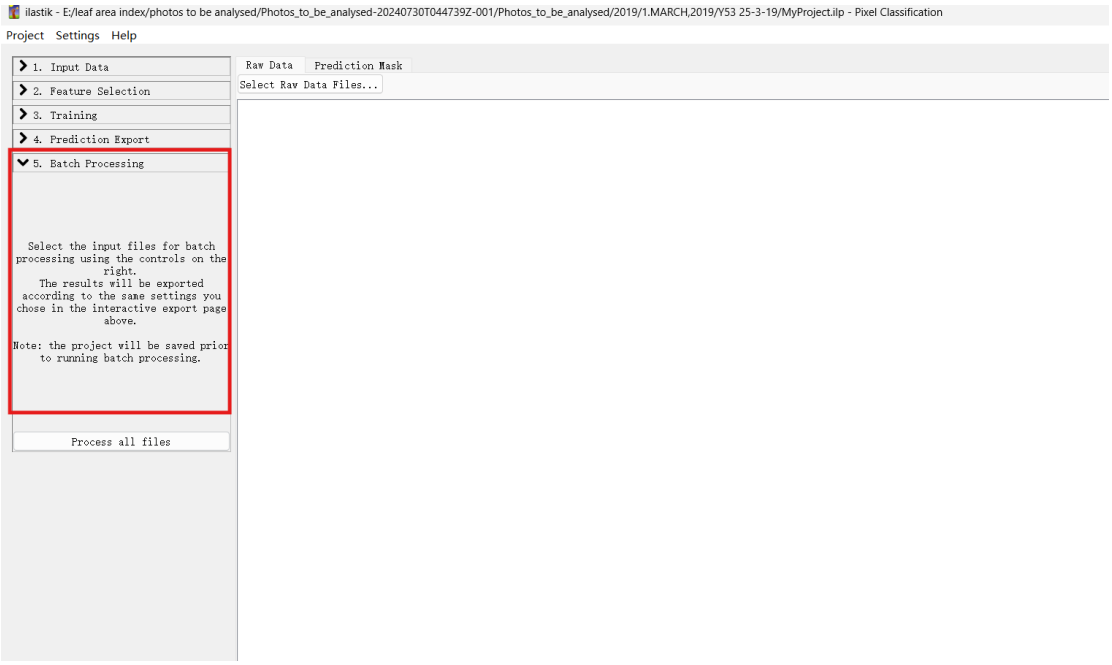


Figure 29

Click 'Select Raw Data Files' and find the folder containing all images you want to analyze (Figure 30). Select all of them and click 'Process All Files'. Use CTRL+A to select all files quickly (Figure 31).

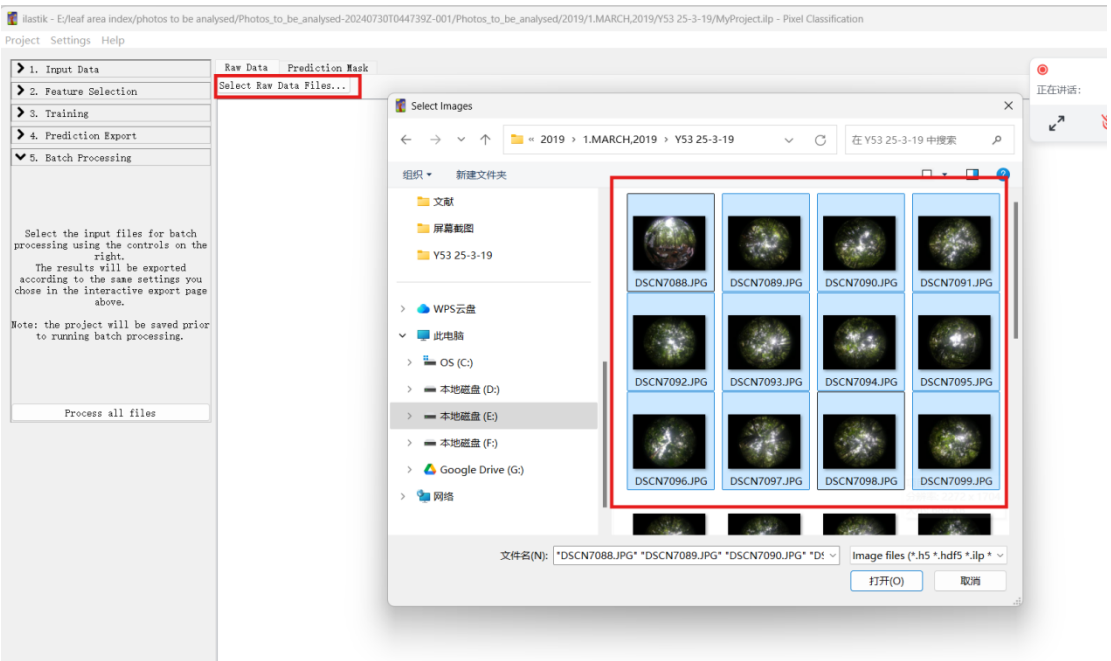


Figure 30

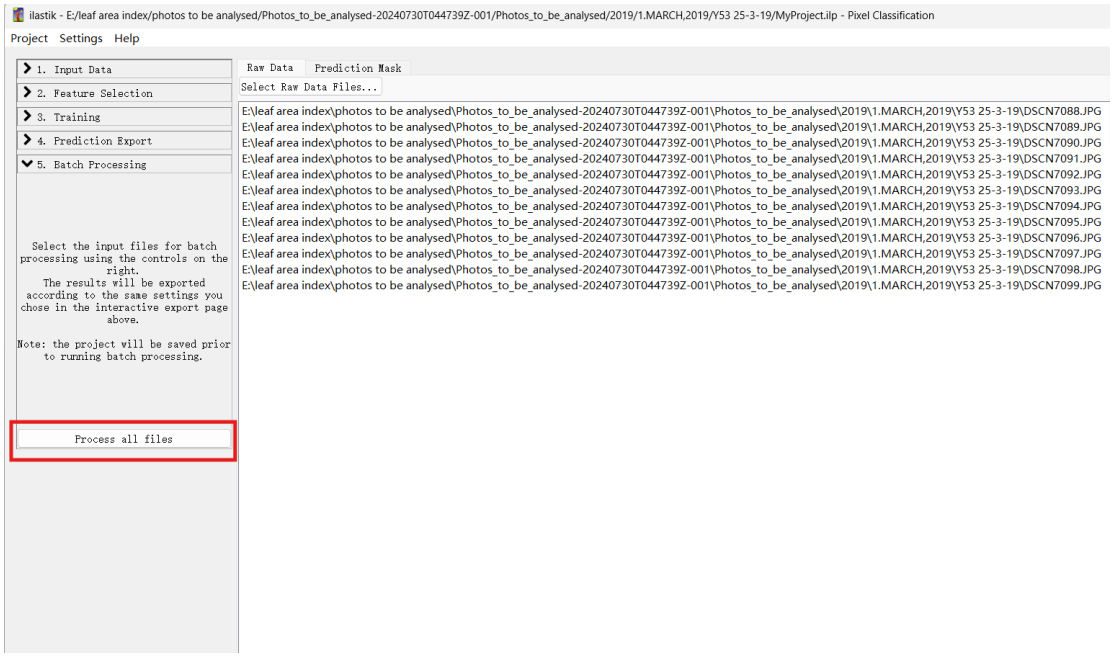


Figure 31

Your output files will be in the output folder you set earlier. For me, it is: C:/Outputs/{nickname}_{result_type}.jpg

Please note that if you save the ilastik project and come back later to batch process more images, **<u>you will have to redo Step 4 (Prediction Export) before Batch Processing, because Step 4 settings are not saved in the .ilp file.</u>**

Using CAN-EYE to process Binary DHP images that have already been classified using ilastik

First, convert all images into black-and-white, or black-and-grey using the machine learning-based software ilastik (as described in the steps above). Then, organize all the monthly and site-specific processed photos in a dedicated folder for each forest (Figure 32).

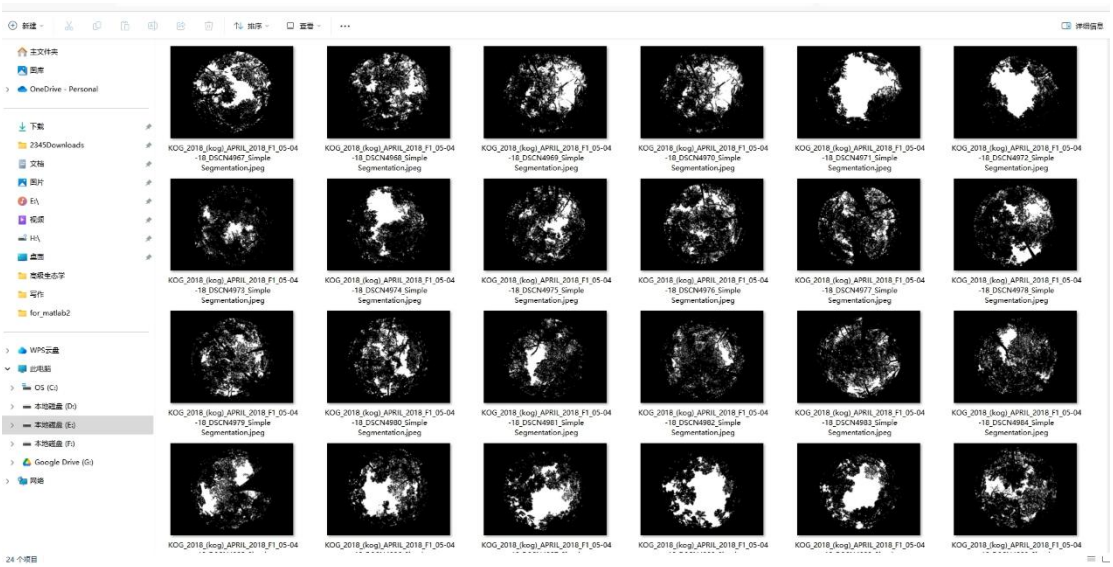
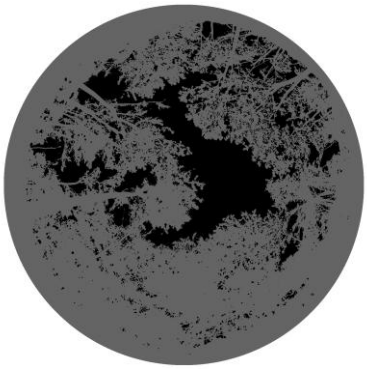


Figure 32

There appears to be a bug in CANEYE V6.49. While the protocol states that only three values are possible—with 0 representing vegetation, 100 indicating a gap, 255 denoting a masked pixel—we discovered that CANEYE actually functions differently: grey (100) represents leaves, while black (0) represents gaps (sky). If you use ilastik with category 1 set as sky and category 2 as leaves, and enable 'Renormalize' to 0-100 option, the output will automatically produce a grey-black image without requiring additional coding. However, if your output is a black-and-white image or entirely black (due to not enabling Renormalize), you will need to convert it to a black-grey format using

code. To assist with this, I have uploaded two scripts: **Image_conversion_20220407.m** and **lets_change_values.R**. Additionally, we mask the area outside the fisheye circle with white (255) for visual clarity, though this step is optional since CAN-EYE handles masking later. Below is a comparison of the images before and after conversion.

Before conversion



After conversion

Important Note: Only a black (sky) and grey (leaves) image is the correct format for CAN-EYE, which differs from the CAN-EYE manual. I believe there is a bug in the program…

Compress the binary images into a zip folder as required by CAN-EYE (Figure 33). For my field setting, we have 14 one-hectare forest plots. Each plot has 25 image-taking pods. Photos are taken every month. Each zip folder contains 25 images, representing one forest plot for a given month. Each zip folder will yield one LAI value.

According to my experience, I can delete up to 15 images in each zip folder. So, if you find any photos that are problematic (i.e. extreme weather, wrong exposure, obstruction by ferns, camera malfunctions, etc), I would suggest deleting them. I don't want to include this in the main text without proper testing, but my impression is that, for dense tropical forests, 15 images per plot should be enough. Using only 10 images per plot is definitely insufficient because some of the photos will inevitably fail due to various unavoidable issues, as always.

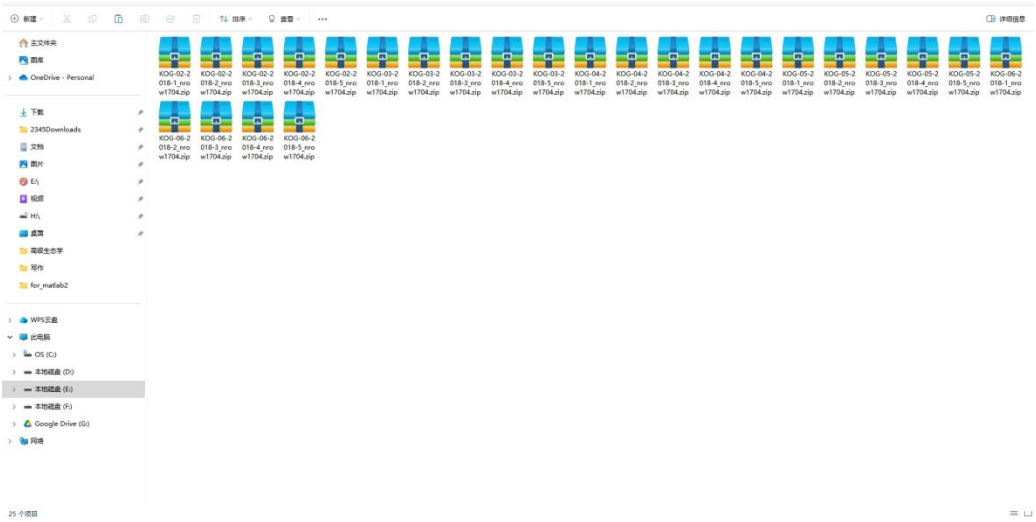


Figure 33

If your CAN-EYE application could not be launched from the desktop, try launching it from this location as an administrator: C:\Program Files\UMT_CAPTE\CAN_EYE\application

In CAN-EYE, we select from the menu: DHP > Binary DHP (Figure 34).

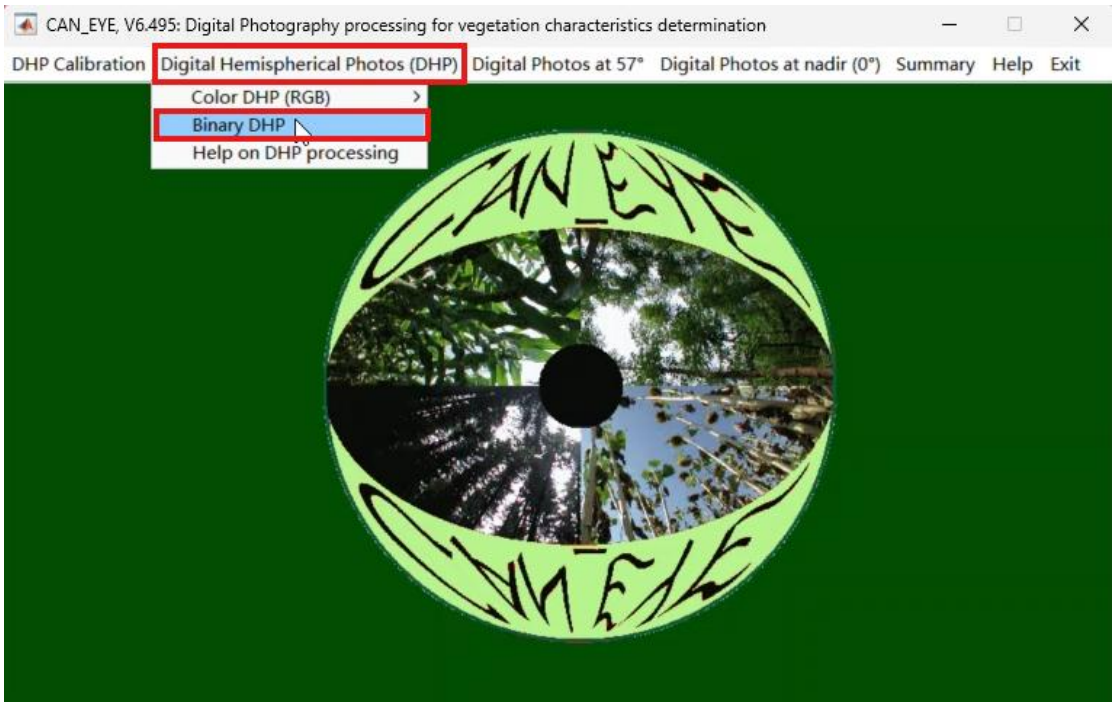
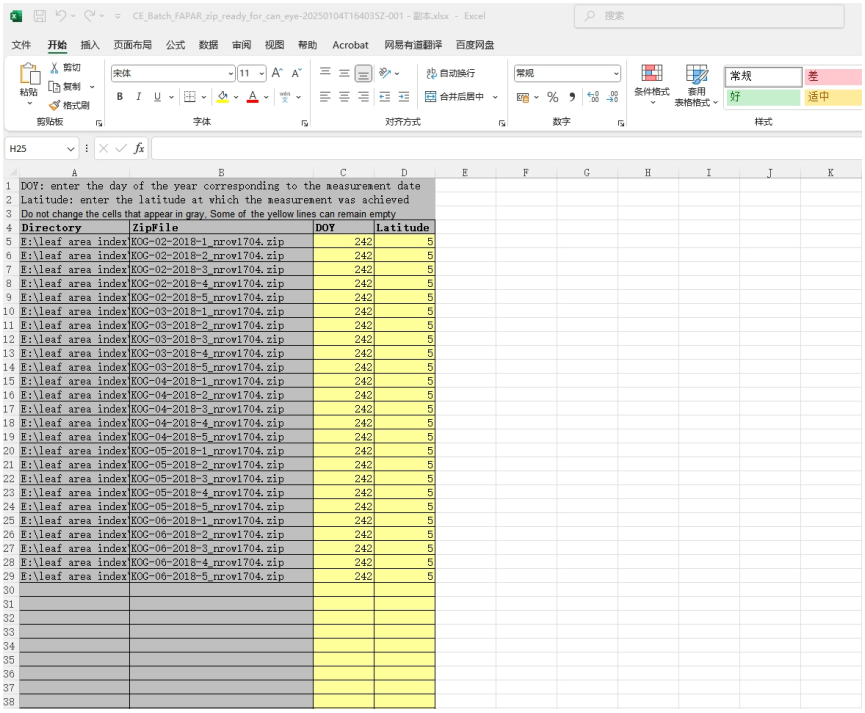


Figure 34

Now, in the Excel file that pops up, we fill in the Day of Year (DOY) and latitude, save the Excel File (Figure 35), and then click 'OK' (Figure 36). The DOY and latitude values will be used to calculate the total daylight hours for fapar estimates, but they are irrelevant to LAI calculations. However, since Kogyae Strict Nature Reserve (my study

site) is very close to the equator, we set the latitude to 5, making DOY unimportant.



DOY: enter the day of the year corresponding to the measurement date
Latitude: enter the latitude at which the measurement was achieved
Do not change the cells that appear in gray. Some of the yellow lines can remain empty

Directory	ZipFile	DOY	Latitude
E:\leaf area index	KOG-02-2018-1_nrov1704.zip	242	5
E:\leaf area index	KOG-02-2018-2_nrov1704.zip	242	5
E:\leaf area index	KOG-02-2018-3_nrov1704.zip	242	5
E:\leaf area index	KOG-02-2018-4_nrov1704.zip	242	5
E:\leaf area index	KOG-02-2018-5_nrov1704.zip	242	5
E:\leaf area index	KOG-03-2018-1_nrov1704.zip	242	5
E:\leaf area index	KOG-03-2018-2_nrov1704.zip	242	5
E:\leaf area index	KOG-03-2018-3_nrov1704.zip	242	5
E:\leaf area index	KOG-03-2018-4_nrov1704.zip	242	5
E:\leaf area index	KOG-03-2018-5_nrov1704.zip	242	5
E:\leaf area index	KOG-04-2018-1_nrov1704.zip	242	5
E:\leaf area index	KOG-04-2018-2_nrov1704.zip	242	5
E:\leaf area index	KOG-04-2018-3_nrov1704.zip	242	5
E:\leaf area index	KOG-04-2018-4_nrov1704.zip	242	5
E:\leaf area index	KOG-04-2018-5_nrov1704.zip	242	5
E:\leaf area index	KOG-05-2018-1_nrov1704.zip	242	5
E:\leaf area index	KOG-05-2018-2_nrov1704.zip	242	5
E:\leaf area index	KOG-05-2018-3_nrov1704.zip	242	5
E:\leaf area index	KOG-05-2018-4_nrov1704.zip	242	5
E:\leaf area index	KOG-05-2018-5_nrov1704.zip	242	5
E:\leaf area index	KOG-06-2018-1_nrov1704.zip	242	5
E:\leaf area index	KOG-06-2018-2_nrov1704.zip	242	5
E:\leaf area index	KOG-06-2018-3_nrov1704.zip	242	5
E:\leaf area index	KOG-06-2018-4_nrov1704.zip	242	5
E:\leaf area index	KOG-06-2018-5_nrov1704.zip	242	5

Figure 35

Figure 36

In the pop-up 'Processing binary images' window, click 'Yes' (Figure 37). We need to set the relevant parameters.

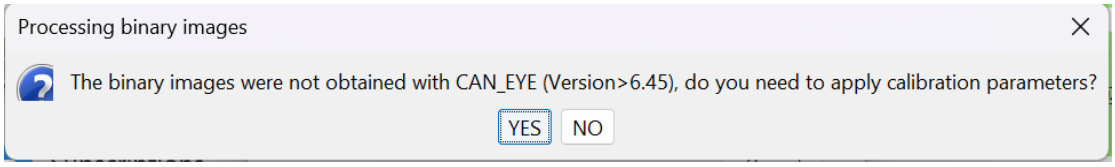


Figure 37

Set the following parameters (Figure 38).

(1) Image Size: Row: 1704, Column: 2272, This could be found by right clicking the image and choose properties (Figure 40)

(2) COI (°) = 80, (field of view is 80 degrees; we exclude the edges as they appear slightly fuzzy)

(3) Sub Sample Factor = 1

(4) Fcover (°) = 20 degrees (calculates the percentage of leaf pixels within the central 20-degree ring, indicating canopy openness. Note: This does not include

the leaf pixels from the entire image).

(5) PAISat = 10 (When a pixel is completely black, the leaf area index (LAI) becomes mathematically infinite. Since CAN-EYE uses 25 subplot images for LAI estimation, this value addresses cases where all subplot shows black at a pixel. We use 10 for dense tropical forests—based on the assumption that the maximum LAI in such forests in 10—and 6 for savannas).

(6) Latitude and Day of Year (Unclear why the software requests DOY again)

Then, click 'Create' to set the parameters for projection and optical center (Figure 39).

(7) Optical center: Enter line 852 and Column 1136 (the center of the fisheye circle), this is half of the 'image size'

(8) P1 = 90/809 = 0.11119, where:

- 90 is the max zenith angle of the fish eye (Note: this example uses a Sigma circular fisheye lens (4.5 mm) with 180 degrees field of view). If you have a 210 degrees field of view fisheye, you will replace 90 with 210/2=105.
- 809 pixels is the distance from the image center to the edge of the fisheye circle (Figure 41). So at 90 degree zenith angle, the image appear at 809 pixels. You can estimate this by using a ruler on your screen to determine how many % of the fish eye edge is completely black, in our case 5%. So we have distance from the center to the edge as 1702/2 * (100 - 5%) = 809. If you want a precise number, you will need software like QGIS or Photoshop.
- Press Enter display the line.

Important note: Fisheye lenses are not linear. You could follow CAN-EYE's more complicated method (via the menu) to derive a curve, but if you prefer simplicity, you can ignore P2 and P3— they have minimal impact on LAI results. Finally, click 'OK'.

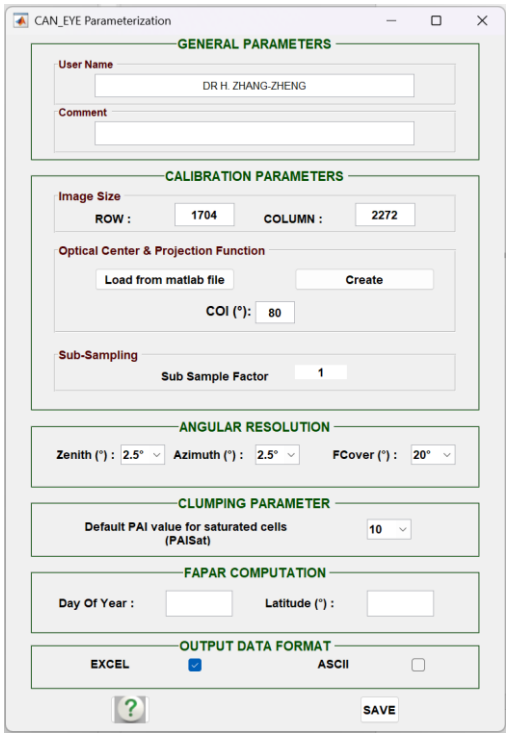


Figure 38

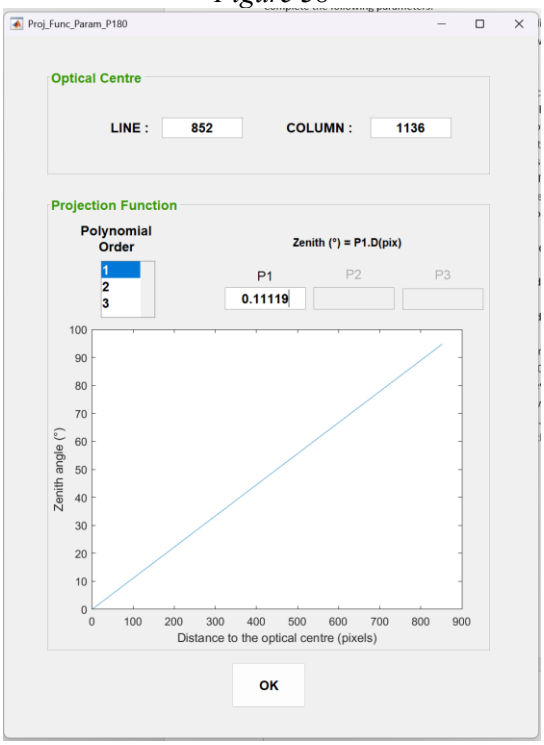


Figure 39

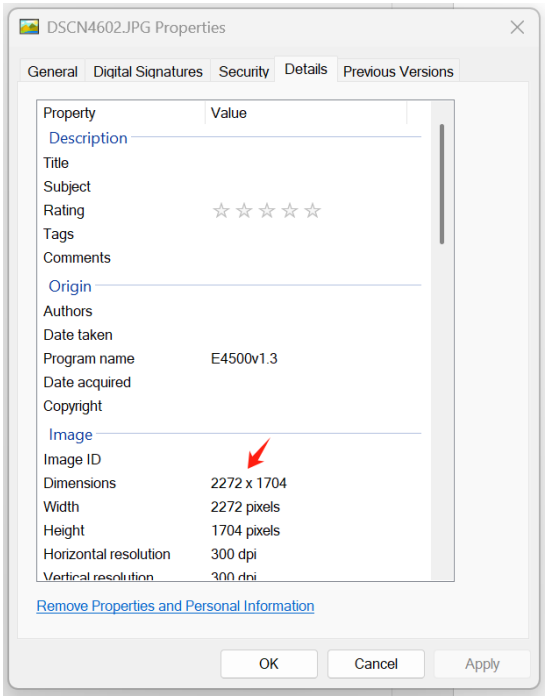


Figure 40

<u>Note: Please do not use the parameter values shown in the screenshots above. You will need to determine your own values for all measurements and calculations (Figure 40 and Figure 41).</u>

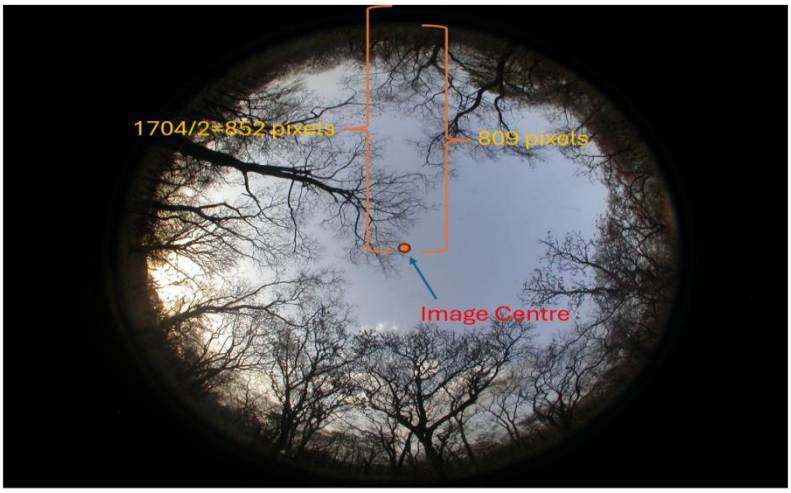


*Figure 41 Number on this screenshot is explained below. 1704 is the height of the image (Figure 40). 852 is the 'line' (Figure 39). 809 is the portion of the 852 without the black edge. In our case, ,we estimated it by 852*0.95. It may not be 0.95 for you. This 809 is used to calculate P1 (Figure 39). 809 represents where the horizon is (because this is 90 degree fish eye lense). You will need to address number 90 based on your fisheye lens. You should adjust the 809 depending on your image height and the thickness of the black edge.*

After clicking 'OK' and wait for the progress bar to finish loading (Figure 42). Next, save the parameter file, which can be directly applied to the remaining zip packages (Figure 43). By default, the parameter files are saved in the following location:

C:\Users\Dr H. Zhang-Zheng\AppData\Local\CAN_EYE\Param_V6

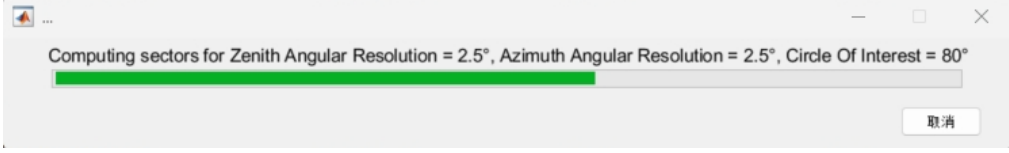


Figure 42

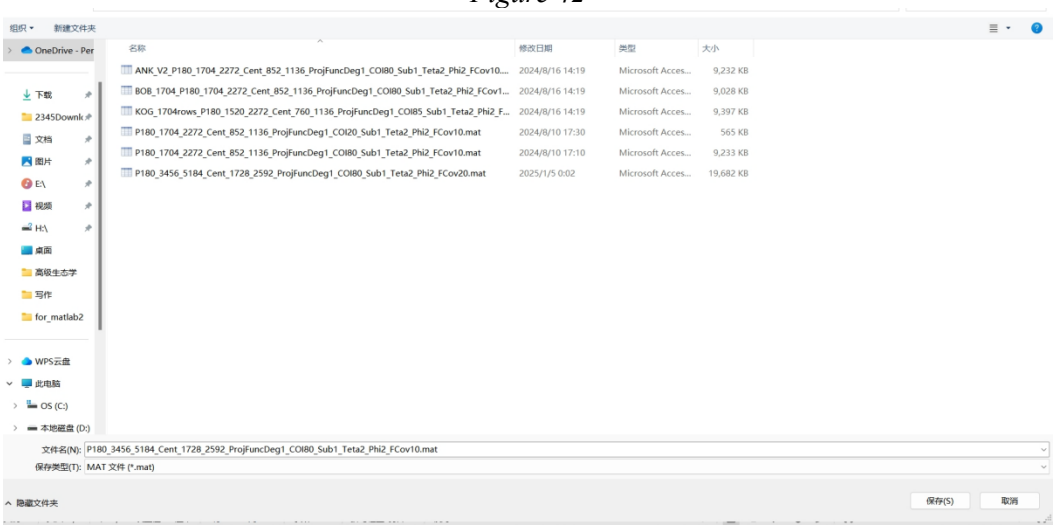


Figure 43

Click 'OK' and wait for CAN-EYE to process the first parcel (Figure 44 and Figure 45).

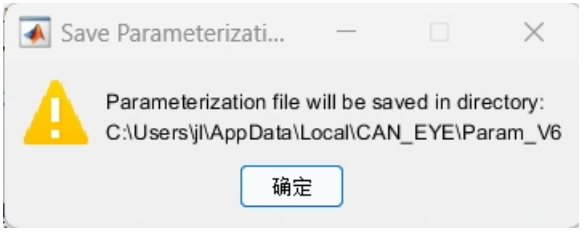


Figure 44

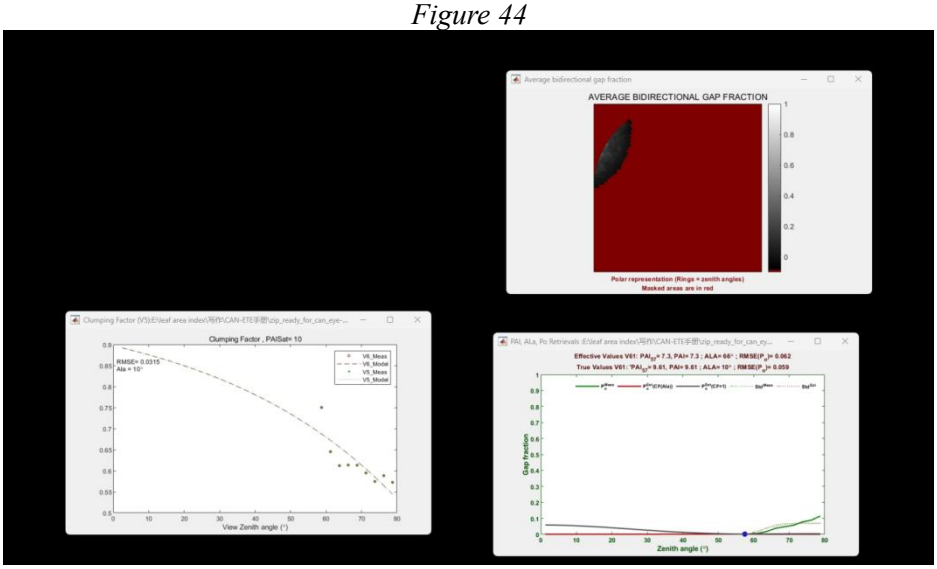


Figure 45

Then, in the pop-up 'Processing binary images' window, click 'YES' to process the second zip folder (Figure 46). Select the previously saved parameter file, click 'Load,'

and wait for CAN-EYE to process the second parcel (Figure 47).

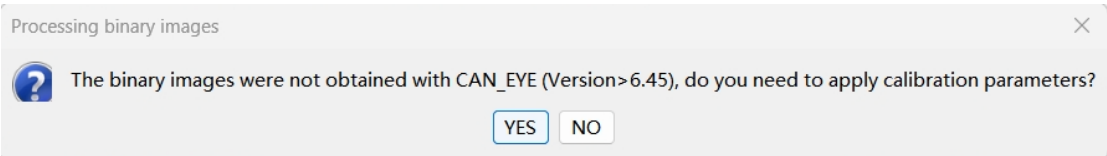


Figure 46

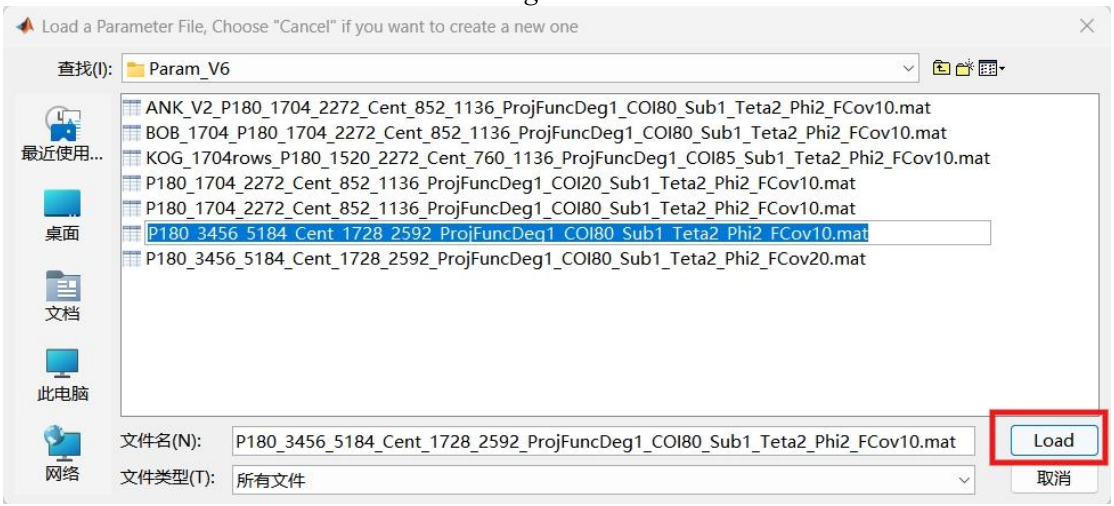


Figure 47

Follow the above process to handle all the remaining parcels.

The output from CANEYE (fAPAR and LAI) consists of an Excel file and an HTML file. You can find the LAI values in these files (Figure 49). We have also prepared R code to read multiple Excel files automatically and join these LAI values into one table. Please find **Gather_LAI_fapar_from_caneye.R**

At this step, if you have lots of image packages, you may have to click 'Load' and 'OK' repeatedly for hours. If you wish to leave your computer unattended to automatically deal with hundreds of packages, you will need a 'mouse clicker app' to help you automatically click this button. We used 'Auto Mouse Click v95.1', but there are many alternative free software.

Leaf area index uncertainty

This is the output from CAN-EYE (Figure 48). Note that for each group of images (25 images in our case, representing 25 subplots in each one-hectare forest plot), the results are generated.

58

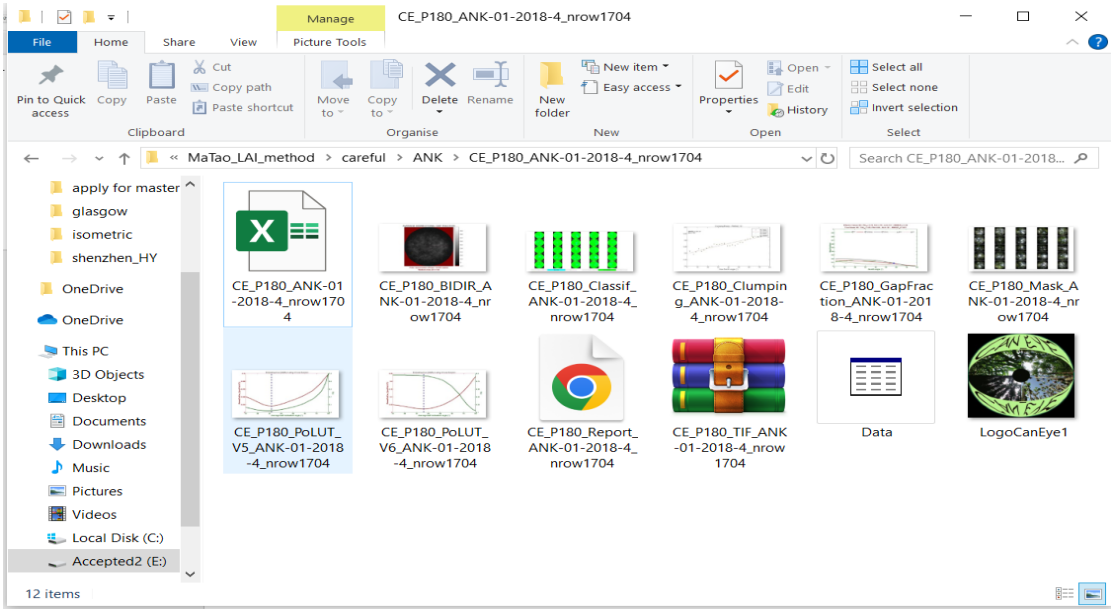
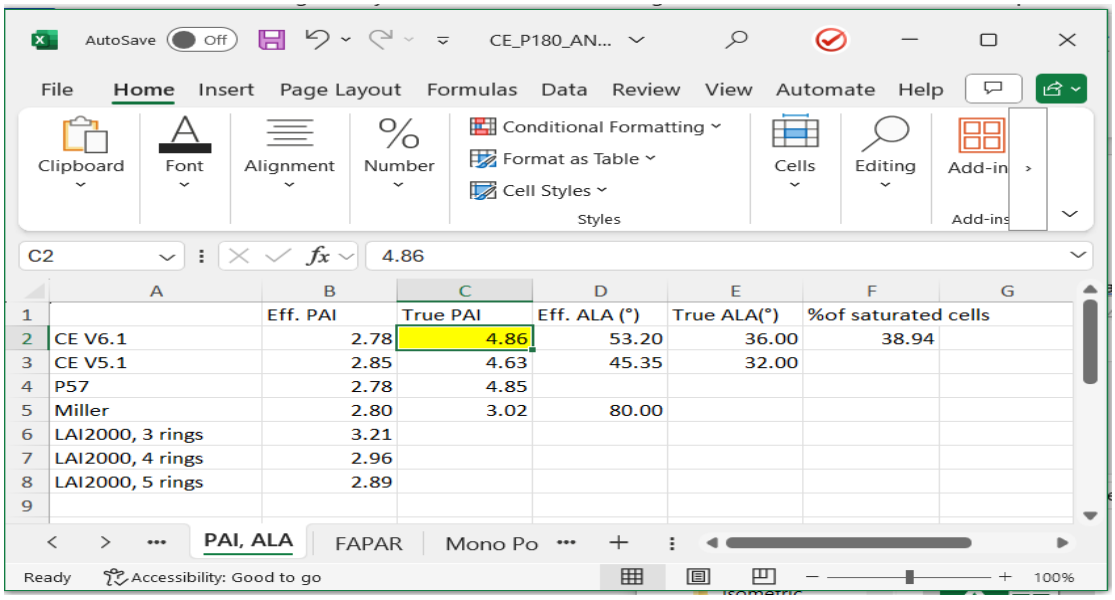


Figure 48

The HTML file shows LAI and fAPAR values in a clear way, but we will extract LAI values from the excel file.

5



	A	B	C	D	E	F	G
1		Eff. PAI	True PAI	Eff. ALA (°)	True ALA(°)	%of saturated cells	
2	CE V6.1	2.78	4.86	53.20	36.00	38.94	
3	CE V5.1	2.85	4.63	45.35	32.00		
4	P57	2.78	4.85				
5	Miller	2.80	3.02	80.00			
6	LAI2000, 3 rings	3.21					
7	LAI2000, 4 rings	2.96					
8	LAI2000, 5 rings	2.89					
9							

Figure 49

We recommend using TRUE PAI under CE V6.1 in tab 'PAI, ALA', which is 4.86 (Figure 49). PAI refers to plant area index. This software does not report leaf area index because it believes that all the hemispherical photos could contain stem and branches.

'True' means the value has been corrected for clumping. PAI (Plant Area Index) is used because your hemispherical images include branches in addition to leaves. For other values, please check CAN-EYE manual.

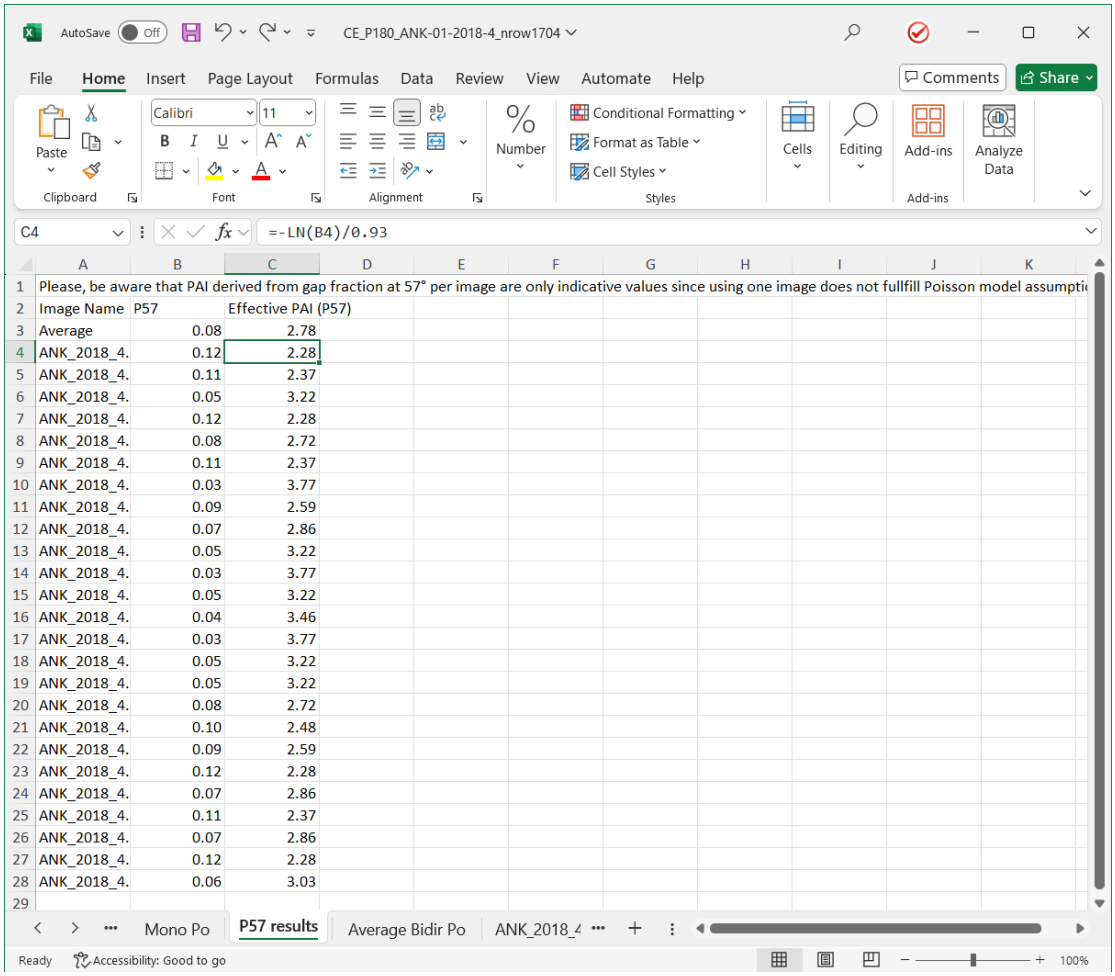
You may also want the 'standard error' for LAI. Unfortunately, CAN-EYE does not provide an LAI value for each image. Such information is likely impossible for CE V6.1, which combines multiple images together. If you use Hemisfer instead of CANEYE, you can get an LAI for each image.

Nonetheless, we can estimate LAI for each image in CANEYE via an indirect approach. Under the 'P57 results' tab, you find P57 values for each image, but not LAI (Figure 50). Note that, as documented in CAN-EYE manual, LAI can be calculated using:

LAI = - LN(P57)/0.93

Then you can obtain LAI for each image following the P57 method. As shown in cell D3, the standard error for P57-based LAI could be calculated (Figure 51). Since we are using CE V6.1, we should scale the standard error from P57 to CE V6.1 (Figure 52), as shown in cell E3. Thus, error_CE_V61 = error_P57 / LAI_P57 * LAI_CE_V61.

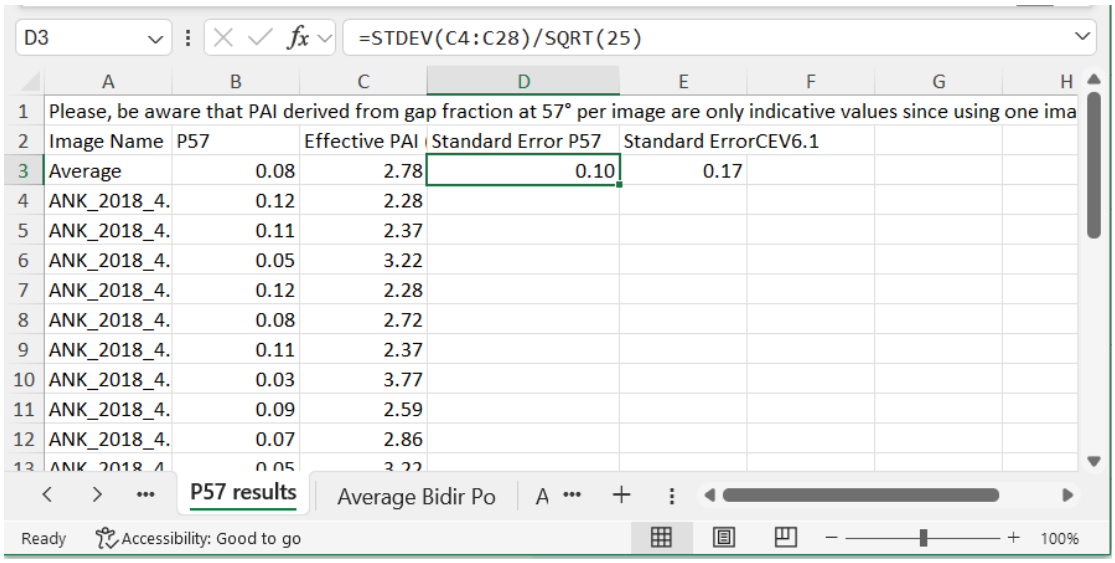
Note that '25' refers to 25 images as listed in Column A. '4.86' refers to the CE V6.1 True PAI. You should use your own value instead of these.



C4 =-LN(B4)/0.93

	A	B	C
1	Please, be aware that PAI derived from gap fraction at 57° per image are only indicative values since using one image does not fullfill Poisson model assumpti		
2	Image Name	P57	Effective PAI (P57)
3	Average	0.08	2.78
4	ANK_2018_4.	0.12	2.28
5	ANK_2018_4.	0.11	2.37
6	ANK_2018_4.	0.05	3.22
7	ANK_2018_4.	0.12	2.28
8	ANK_2018_4.	0.08	2.72
9	ANK_2018_4.	0.11	2.37
10	ANK_2018_4.	0.03	3.77
11	ANK_2018_4.	0.09	2.59
12	ANK_2018_4.	0.07	2.86
13	ANK_2018_4.	0.05	3.22
14	ANK_2018_4.	0.03	3.77
15	ANK_2018_4.	0.05	3.22
16	ANK_2018_4.	0.04	3.46
17	ANK_2018_4.	0.03	3.77
18	ANK_2018_4.	0.05	3.22
19	ANK_2018_4.	0.05	3.22
20	ANK_2018_4.	0.08	2.72
21	ANK_2018_4.	0.10	2.48
22	ANK_2018_4.	0.09	2.59
23	ANK_2018_4.	0.12	2.28
24	ANK_2018_4.	0.07	2.86
25	ANK_2018_4.	0.11	2.37
26	ANK_2018_4.	0.07	2.86
27	ANK_2018_4.	0.12	2.28
28	ANK_2018_4.	0.06	3.03
29			

Figure 50



D3 =STDEV(C4:C28)/SQRT(25)

	A	B	C	D	E
1	Please, be aware that PAI derived from gap fraction at 57° per image are only indicative values since using one ima				
2	Image Name	P57	Effective PAI	Standard Error P57	Standard ErrorCEV6.1
3	Average	0.08	2.78	0.10	0.17
4	ANK_2018_4.	0.12	2.28		
5	ANK_2018_4.	0.11	2.37		
6	ANK_2018_4.	0.05	3.22		
7	ANK_2018_4.	0.12	2.28		
8	ANK_2018_4.	0.08	2.72		
9	ANK_2018_4.	0.11	2.37		
10	ANK_2018_4.	0.03	3.77		
11	ANK_2018_4.	0.09	2.59		
12	ANK_2018_4.	0.07	2.86		

Figure 51

E3 =D3/C3*4.86

	A	B	C	D	E	F	G	H
1	Please, be aware that PAI derived from gap fraction at 57° per image are only indicative values since using one ima							
2	Image Name	P57	Effective PAI	Standard Error P57	Standard ErrorCEV6.1			
3	Average	0.08	2.78	0.10	0.17			
4	ANK_2018_4.	0.12	2.28					
5	ANK_2018_4.	0.11	2.37					
6	ANK_2018_4.	0.05	3.22					
7	ANK_2018_4.	0.12	2.28					
8	ANK_2018_4.	0.08	2.72					
9	ANK_2018_4.	0.11	2.37					
10	ANK_2018_4.	0.03	3.77					
11	ANK_2018_4.	0.09	2.59					
12	ANK_2018_4.	0.07	2.86					
13	ANK_2018_4.	0.05	3.22					

P57 results | Average Bidir Po | A

Ready Accessibility: Good to go 100%

Figure 52

	A	B	C	D	E	F	G	H	I	J
1	Please, be aware that PAI derived from gap fraction at 57° per image are only indicative values since using one image does not fullfill Poisson m									
2	Image Name	P57	Effective PAI	Standard Error P57	Standard Erro	Systematic	Uncertainty			
3	Average	0.08	2.78	0.10	0.17	0.486	0.52			
4	ANK_2018_4.	0.12	2.28							
5	ANK_2018_4.	0.11	2.37							
6	ANK_2018_4.	0.05	3.22							
7	ANK_2018_4.	0.12	2.28							
8	ANK_2018_4.	0.08	2.72							
9	ANK_2018_4.	0.11	2.37							
10	ANK_2018_4.	0.03	3.77							
11	ANK_2018_4.	0.09	2.59							
12	ANK_2018_4.	0.07	2.86							
13	ANK_2018_4.	0.05	3.22							

P57 results | Average Bidir Po | ANK_2018_4._APF

Figure 53

The standard error accounts for the random errors of associated with hemispherical photo sampling. We have not accounted for systematic errors, which originate from the algorithm used to convert gap fraction to LAI. Here, we set the systematic error at 10% (Figure 53). Thus, systematic_errors = 0.1 * LAI_CE_V61 = 0.486.

To combine both systematic and random errors (Figure 53), the uncertainty is calculated as = sqrt (systematic^2 + random^2) = sqrt(0.17^2 + 0.486^2) = 0.52.

Note that 0.52 is about 10.7% of the LAI value (4.86). This analysis thus shows that systematic error dominates in the leaf area index calculation, in agreement with the literature (see this). Unfortunately, there is no better method for calculating systematic error than an arbitrary estimate.

In short, if you find the above confusing, for simplicity we recommend using an arbitrary 15% uncertainty. For LAI = 4.86, this gives 4.86 ± 4.86*0.15 = 4.86 ± 0.729

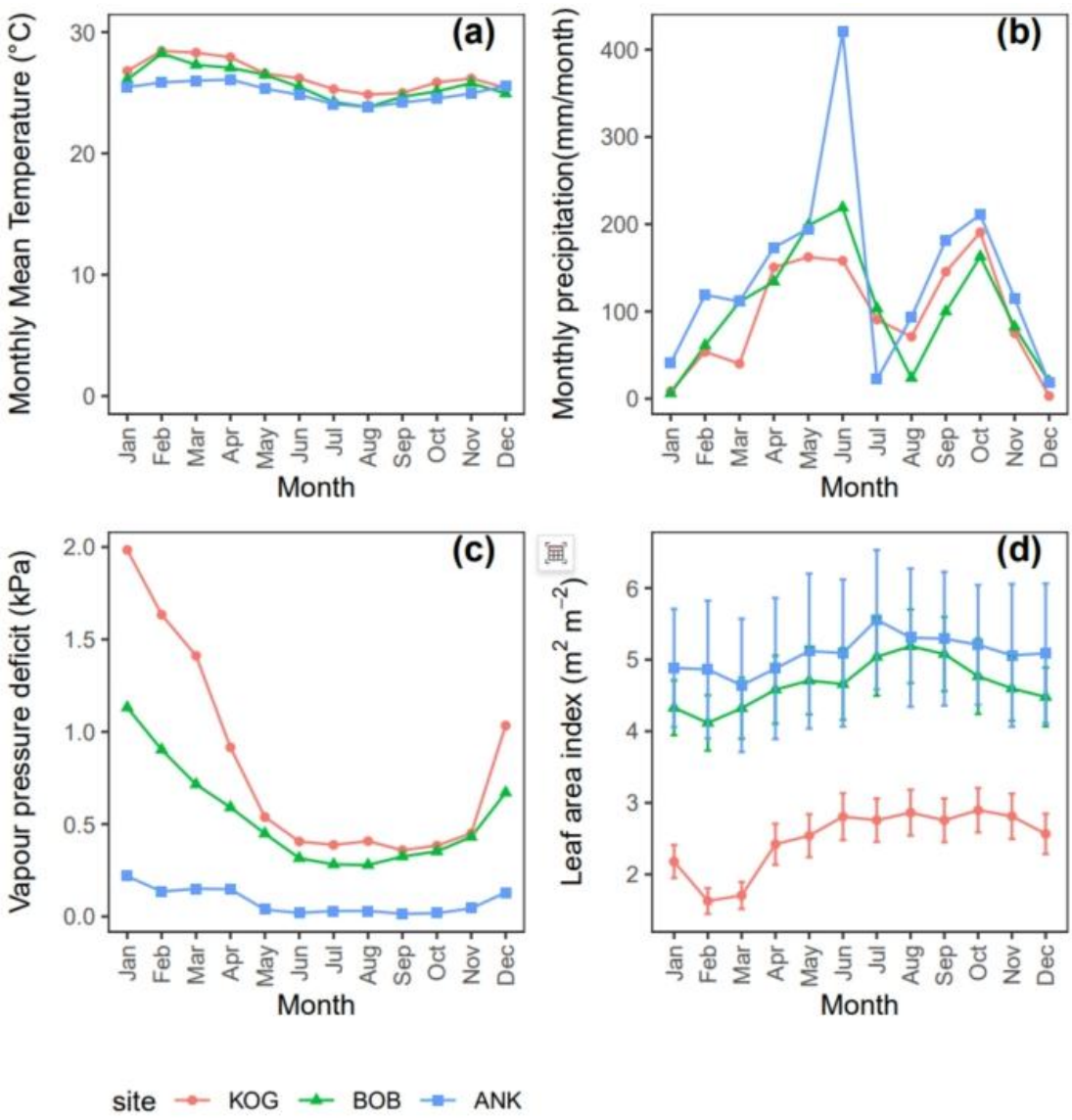
Supplementary information for 'An AI-based and coding-free protocol for forests Leaf Area Index (LAI) calculation'

Tao Ma[1,2], Minxue Tang[3], Fangni Liu[1], Akwasi Duah-Gyamfi[4], Stephen Adu-Bredu[4], Mark Arcebal K. Naive[5,6], Imma Oliveras Menor[2,7], Sam Moore[2], Zhiyuan Zhang[8], Yadvinder Malhi[2,9]✉, Cecilia A. L. Dahlsjö[2,9], Huanyuan Zhang-Zheng[2,9]✉

✉ Corresponding Authors
(yadvinder.malhi@ouce.ox.ac.uk)
(huanyuan.zhang@ouce.ox.ac.uk)

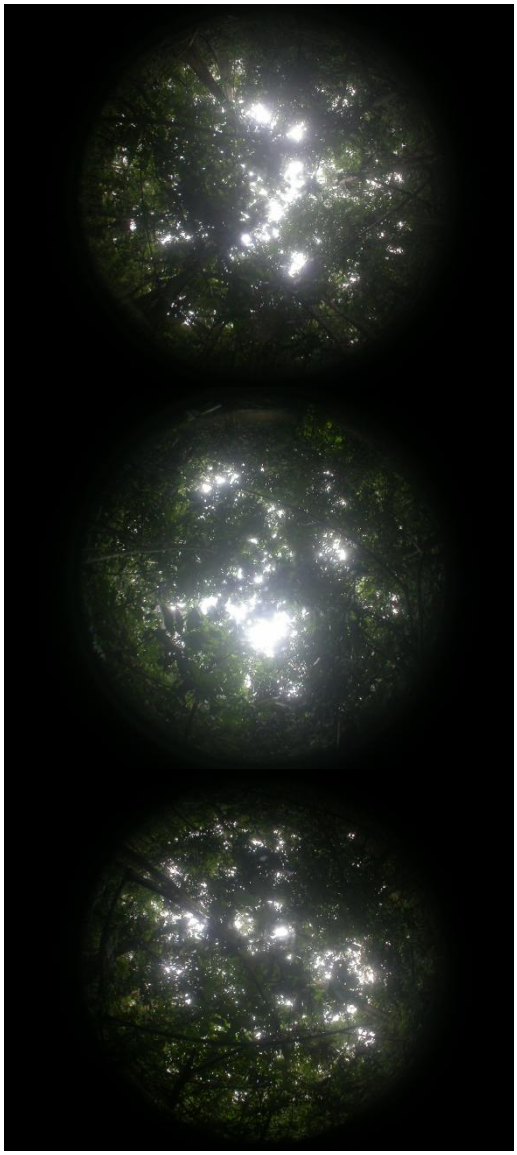
Affiliations

1 College of Resources and Environmental Sciences, China Agricultural University, Beijing, China
2 Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, United Kingdom
3 Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, United Kingdom
4 Forestry Research Institute of Ghana, Council for Scientific and Industrial Research, Kumasi, Ghana
5 Center for Integrative Conservation and Yunnan Key Laboratory for Conservation of Tropical Rainforests & Asian Elephants, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China
6 University of Chinese Academy of Sciences, Beijing 100049, China
7 AMAP – botAnique et Modélisation de l'Architecture des Plantes et des Végétations, Université de Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier CEDEX 5, France
8 College of Resources and Environment, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China
9 Leverhulme Centre for Nature Recovery, University of Oxford, Oxford, United Kingdom

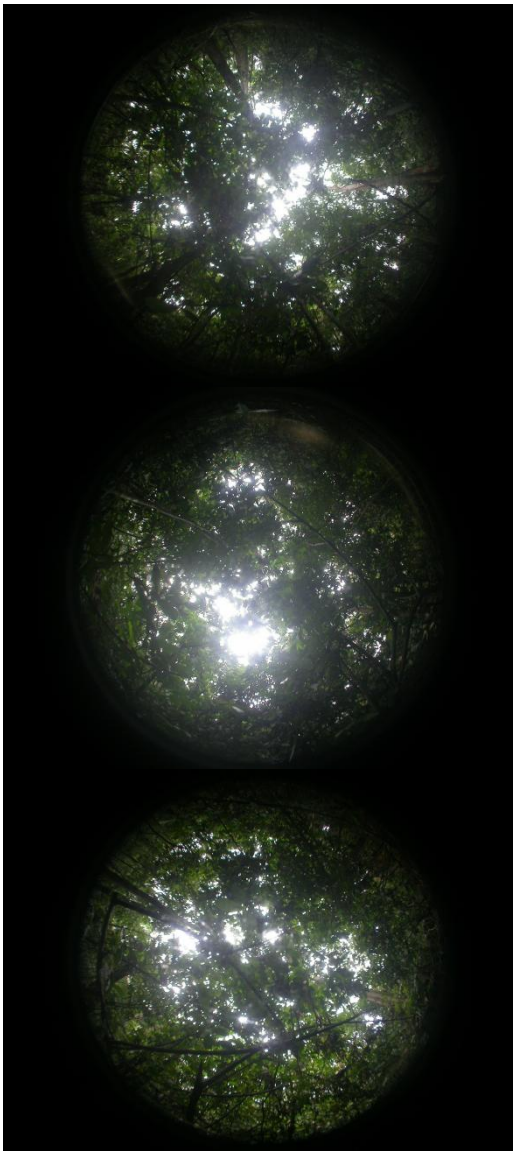


Supplementary Figure 1 Study site seasonality. We show (a) Monthly mean temperature (oC) (b) monthly sum precipitation (mm/month) (c) vapour pressure deficit (kPa) and (d) leaf area index (m2 m-2) for the three study sites. Monthly mean LAI was a mean across 2016-2018(for ANK) and 2012-2019 (for BOB&KOG). Climate data were reported by local meteological stations across 2011(for ANK), 2012-2016 (for BOB), and 2013-2016(for KOG)

ANK-01　2018.7

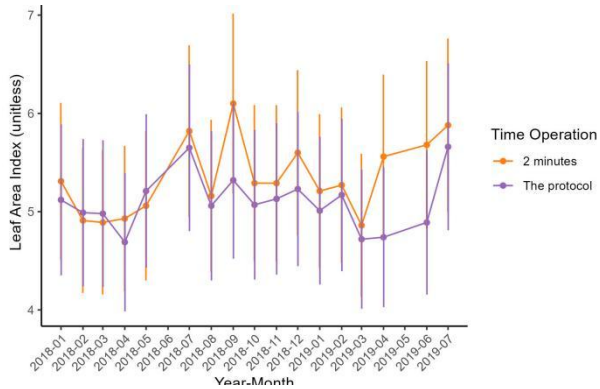


ANK-01　2018.9



Supplementary Figure 2　Fish-eye photos (ANK-01) taken on 2018 Sept and July. By comparing with the naked eye, we can clearly see that in the fish-eye photo from September 2018 on the right, there are fewer leaves because its sky window is larger and the edges are more fragmented.

(A) Monthly LAI Comparison at ANK-01: CAN-EYE (2-Minutes) vs. The protocol



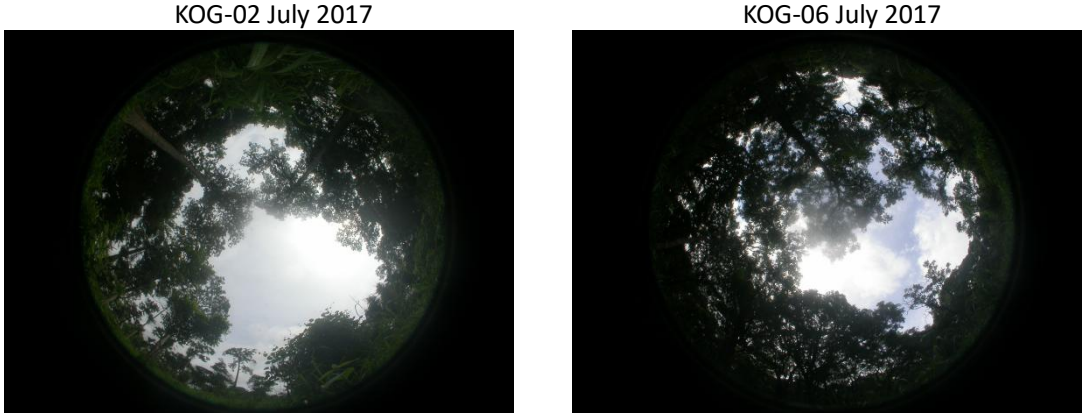
Orange line: CAN-EYE two-minute processing (2 Minutes);
Purple line: ilastik fully automated processing(The protocol);
X-axis: Month;
Y-axis: Leaf Area Index.

(B) ANK-01 2019.2

(C) ANK-01 2019.4

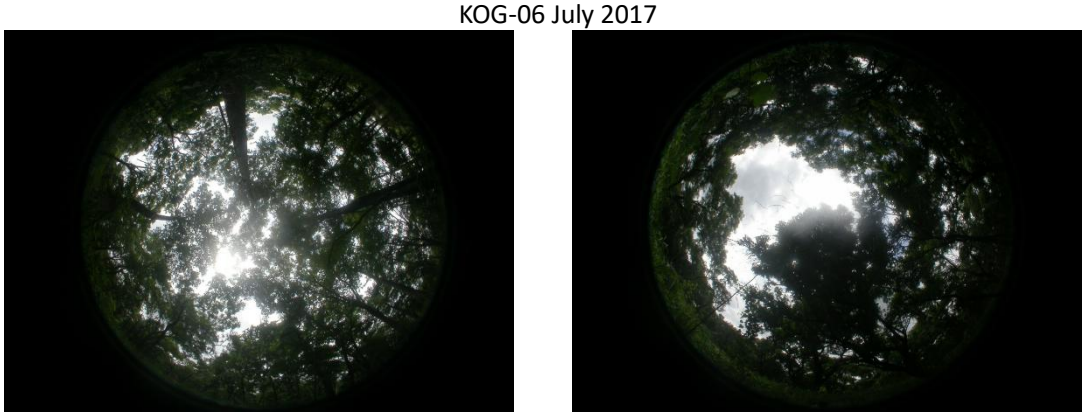
Supplementary Figure 3 Fisheye photos of ANK-01 (proving ilastik fully automated to be more accurate). Ilastik's results show February 2019 is significantly higher than April 2019, while CAN-EYE manual results indicate the opposite (A). However, visual observation clearly shows more leaves in February 2019 (B and C).

KOG-02 July 2017 KOG-06 July 2017

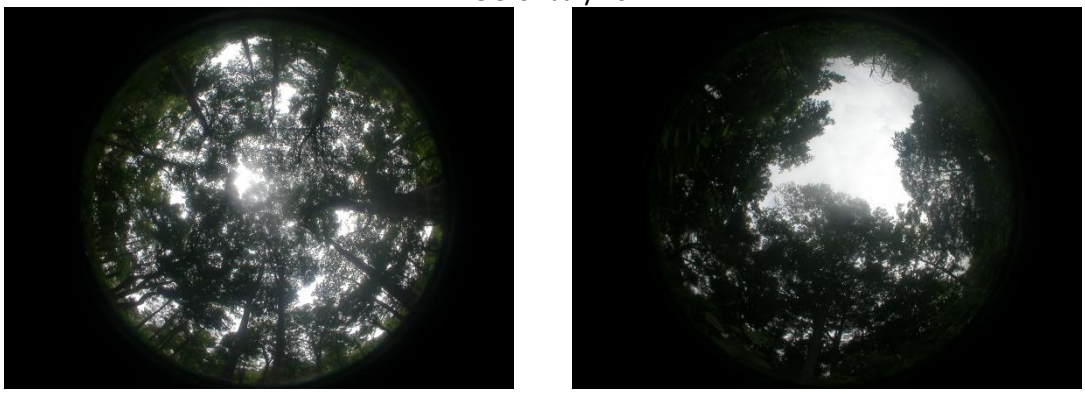


Supplementary Figure 4 July fisheye photos of KOG-02 and KOG-06 (with a large sky window). KOG-02 is a dry forest and KOG-06 is a savanna.

KOG-06 July 2017



KOG-02 July 2017



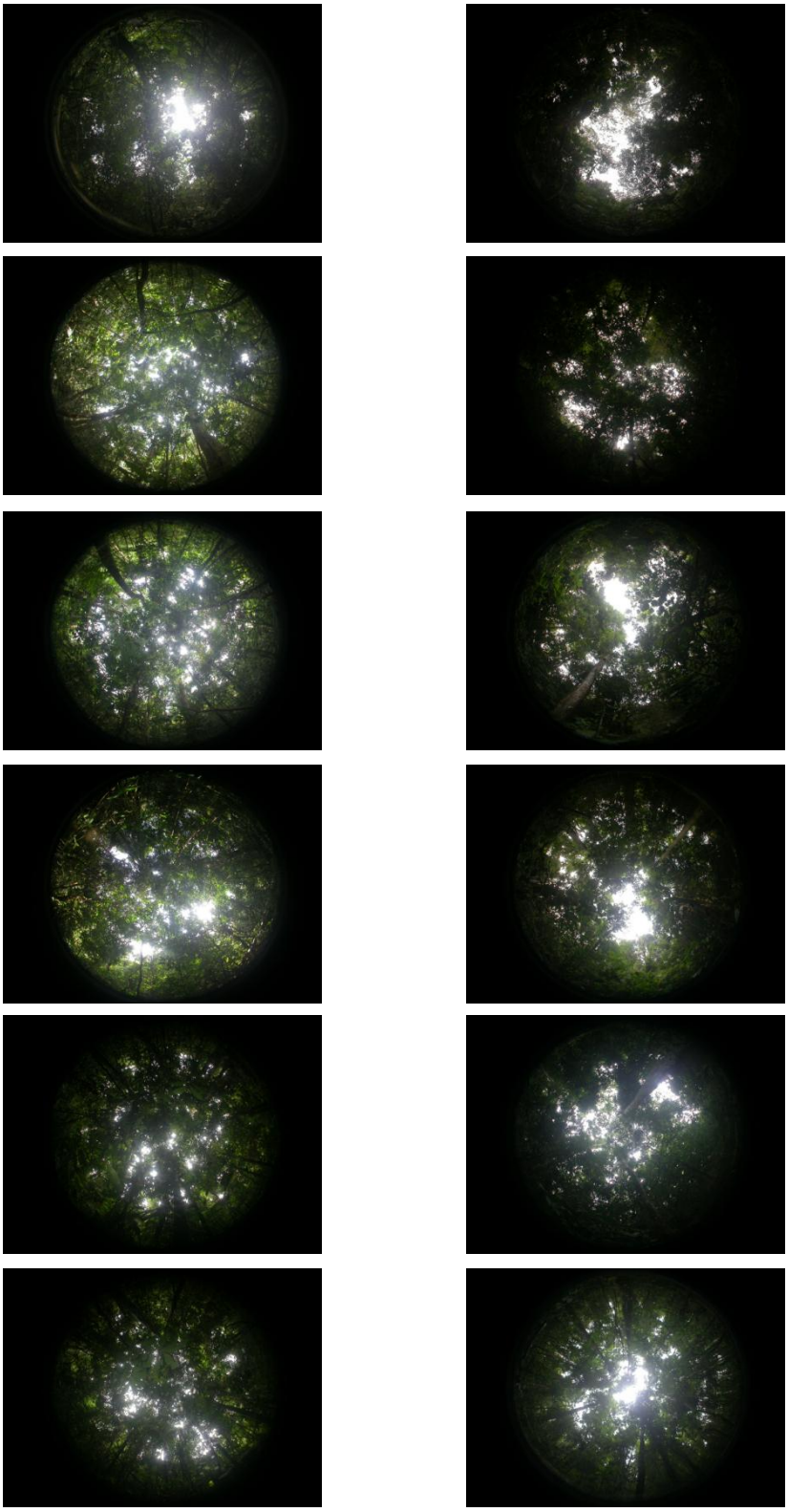
Supplementary Figure 5 July fisheye photos of KOG-02 and KOG-06 (very different from the same site). If the left side of the photo is well marked through CAN-EYE, the right side of the photo in the same batch of photos will put a large area of the sky into the leaves. On the contrary, if the right image is well labeled, many leaves on the left will be classified as the sky.

ANK-03 BOB-04

2018-1

2018-3

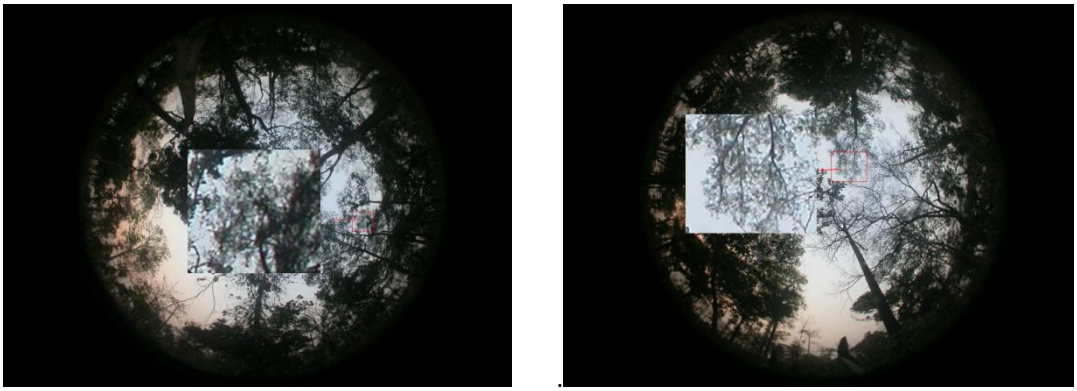
2018-5



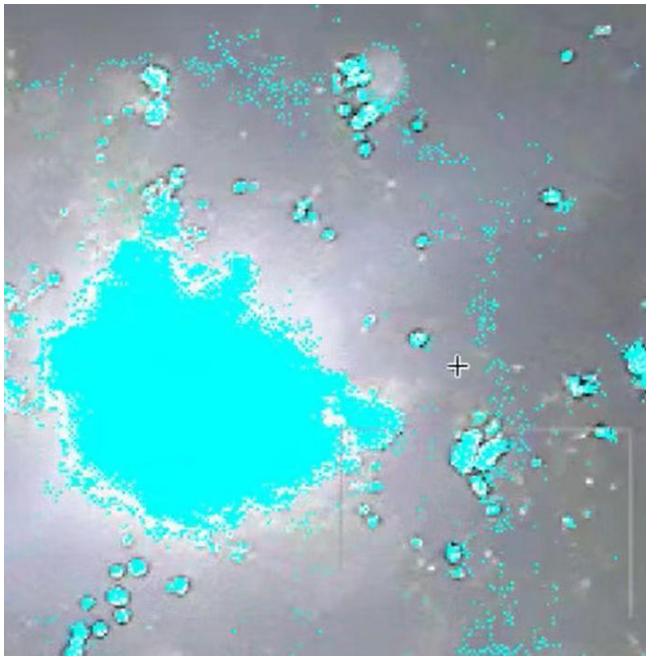
Supplementary Figure 6 Fisheye photos of ANK-03 and BOB-04. Here are the fish-eye photos of six sites in ANK-03 from January, March, and May of 2018, with two sites from each month, totaling six distinct sites. It also shows the six sites in BOB-04 from January, March, and May of 2018, with each site being different. By visually comparing these fish-eye photos, it is apparent that there are more leaves in ANK-03.

Intensive sky beam

Small and dense leaf

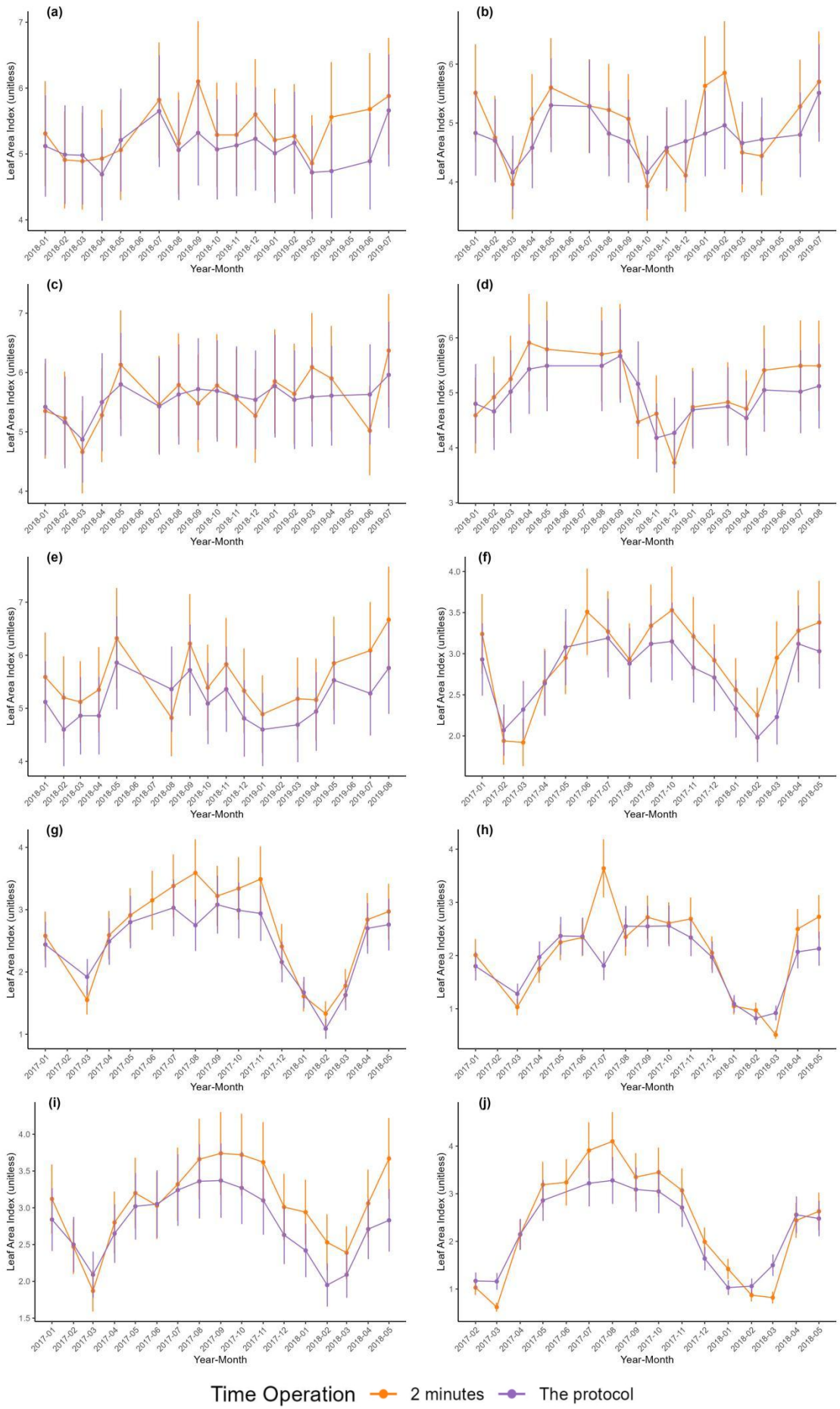


Supplementary Figure 7 The ' gray zone ' on hemispherical photos. Here are some images of ‘gray areas’, here we show two examples: (1) intensive sky beam, sunlight erosion on the leaves, making it difficult for humans to judge. (2) Small and dense leaf distribution, manual classification is hard to operate, the border between leaves and skies are fused and vague. Overall, it is challenging for humans to make decisions when processing these areas. Subjective errors are particular large when dealing with these photos.

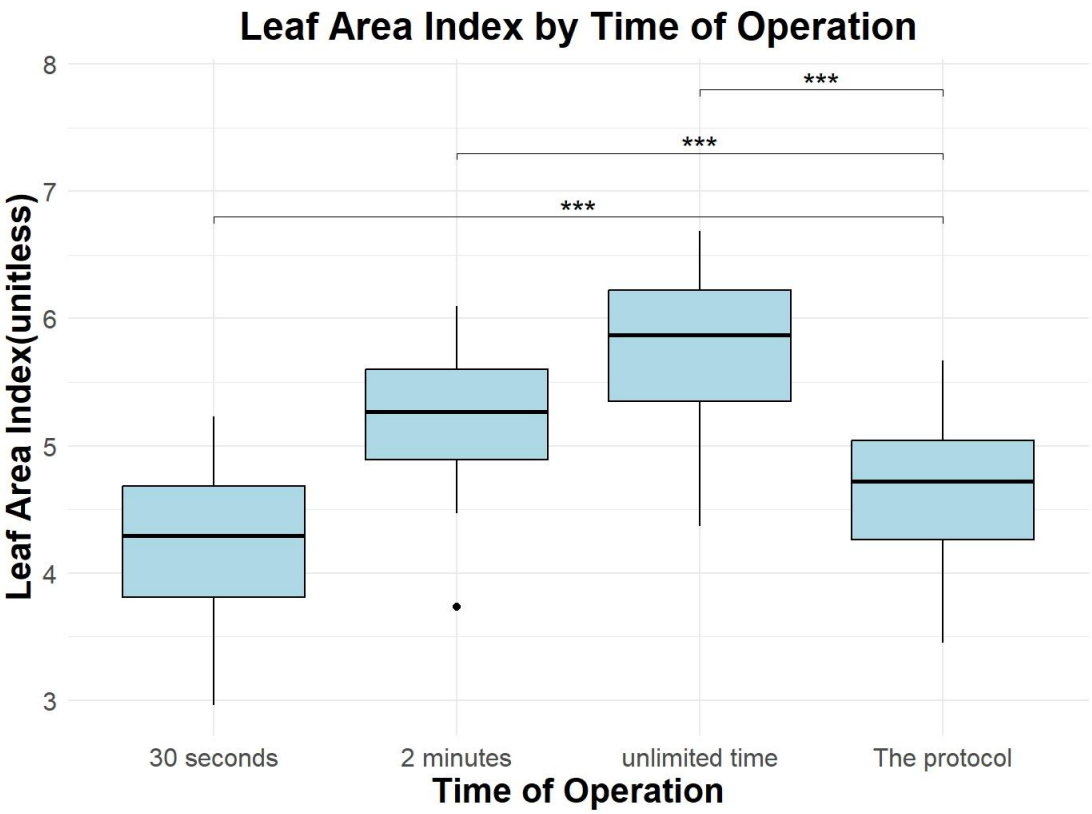


Supplementary Figure 8 Error pixel clusters in CAN-EYE processing. In the edge regions of the sky (especially when sunlight is intense), CAN-EYE often produces dense pixel points, with gaps between them. These gaps are classified as part of the leaves, but in reality, they should be considered part of the sky. This situation is difficult to avoid, even with very careful processing. In this screenshot, blue is the classification outcome of sky. The background is a zoom in of a hemispherical photo.

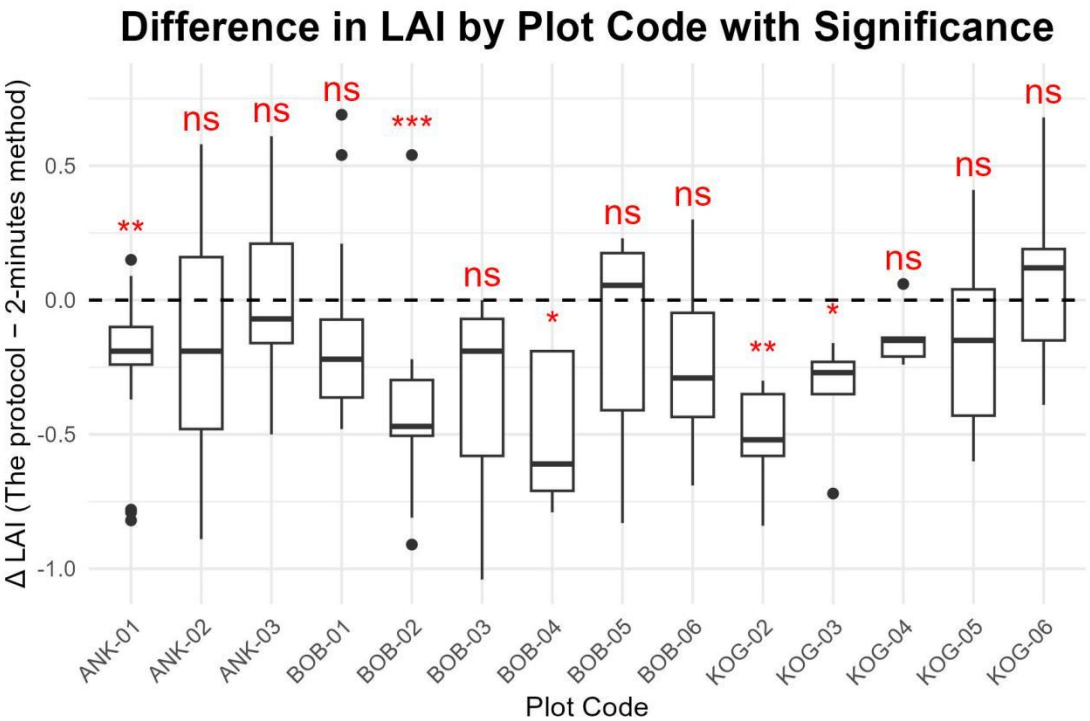
Supplementary Figure 9 Fisheye photos under CAN-EYE binary classification method. In the following scenario, the content within the green box should be classified as leaves, as the leaves are clearly visible on the branches (visible on a 4K monitor displaying the raw image, please zoom in to check details). The content within the red box should be classified as the sky, as this area only contains branches with no leaves. However, in AI-based binary classification, both of these situations are classified as leaves.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
Leaf Area Index (unitless)
Year-Month
Time Operation
2 minutes
The protocol

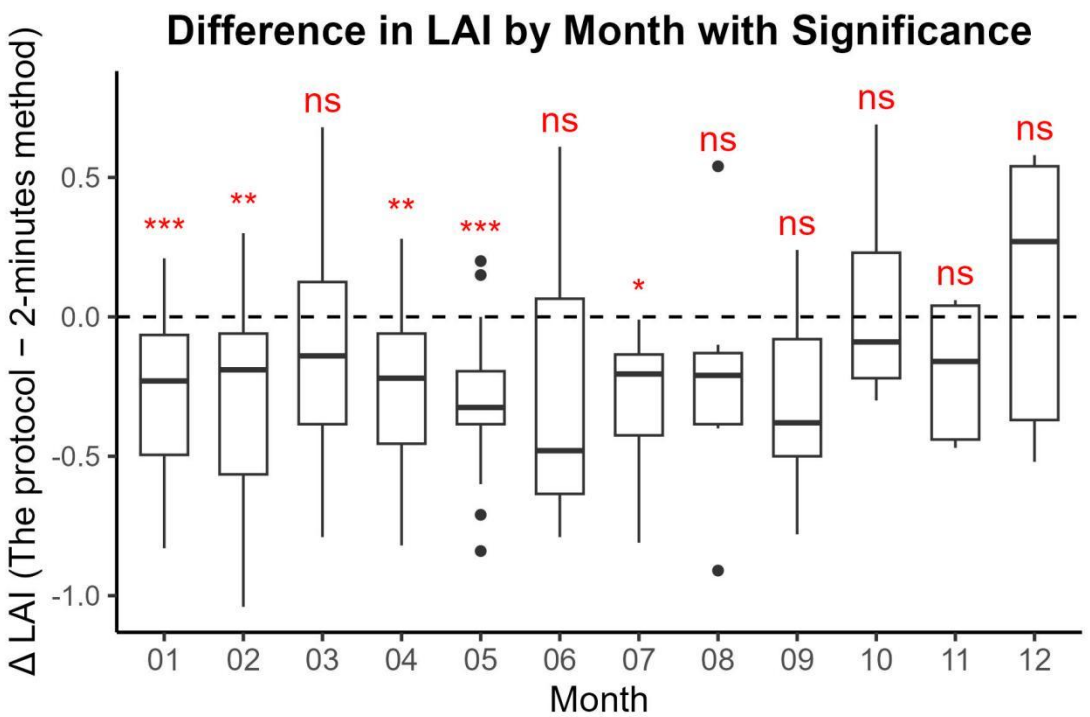
Supplementary Figure 10 Time series of Leaf Area Index (LAI) derived from different processing methods across multiple forest plots.LAI dynamics are shown for each individual plot using two distinct image processing methods. The orange line indicates LAI values processed with the CAN-EYE method, which requires approximately two minutes per image. The purple line represents results generated by the protocol, which requires only a one-time training step, followed by automated batch processing. These comparisons provide insight into the temporal patterns and methodological consistency across different forest types and image acquisition settings.(a) ANK-01; (b) ANK-02; (c) ANK-03; (d) BOB-01; (e) BOB-02; (f) KOG-03; (g) KOG-04; (h) KOG-05; (i) KOG-02; (j) KOG-06. Note that the two-minute CAN-EYE processing experiments used images from the following periods; – **ANK-01**, **ANK-02**, **ANK-03**: photographs from December 2017 to August 2019
– **BOB-01**, **BOB-02**: photographs from January 2018 to September 2019
– **KOG-02**, **KOG-03**, **KOG-04**, **KOG-05**, **KOG-06**: photographs from January 2017 to May 2018. For **BOB-03**, **BOB-04**, **BOB-05**, and **BOB-06**, only images from January to May 2018 were available for the two-minute CAN-EYE experiment, which was insufficient to capture full seasonal dynamics. Therefore, these plots were not included in Supplementary Fig. 12.



Supplementary Figure 11 Comparison of leaf area index (LAI) estimates under four different image processing methods: the protocol, 30 seconds, 2 minutes, and unlimited time. Boxplots show the distribution of LAI values obtained from two forest plots (ANK-01 and BOB-01) for each operation methods. A linear mixed-effects model was used to compare groups with the protocol as the reference and plot identity as a random effect. Asterisks (***) indicate statistically significant differences from the protocol ($p < 0.001$ for all comparisons).



Supplementary Figure 12 Boxplots of **Δ LAI (The protocol minus 2-minutes method)** across individual forest plots. Each box represents the distribution of LAI differences between the two methods within a given plot. Positive values indicate that the protocol method yielded higher LAI estimates than the 2-minutes manual method. Significance levels (*, **, ***) indicate whether the mean Δ LAI significantly differs from zero based on a one-sample t-test ($p < 0.05$, 0.01, 0.001), while "ns" denotes non-significance.



Supplementary Figure 13 Boxplots of **Δ LAI (The protocol minus 2-minutes method)** by calendar month (1 = January…12 = December). Each box depicts the distribution of LAI differences pooled across all plots in a given month. Positive Δ LAI values indicate that the protocol workflow yielded

higher LAI estimates than the 2-minutes manual method. Significance of the mean Δ LAI versus zero was assessed by one-sample t-tests, with symbols denoting $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***); "ns" indicates non-significance.