High-throughput seed phenotyping of Populus cultivars in China using vibration-assisted machine vision with alternating back-lit and front-lit illuminations

Xiwei Wang[1*], Mubikayi Muhong Horly[1], Zanpeng Li[1], Maocheng Zhao[1], Bin Wu[1], Mengmeng Wang[1], Jiale Deng[1], Mengmeng Qiao[1], Xiao Chen[1], Yongjie Lu[1], Liang Qi[1], Weijun Xie[1], Hongyuan Zou[1], Yan Zhang[2], Rusheng Peng[2]

(1. *College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China*;
2. *Liaoning Institute of Poplar Research, Gaizhou 115213, Liaoning, China*)

Abstract: Seeds of major Populus cultivars were collected from across China in 2024 to build the image-data bank of over 1 187 000 images of singular seeds for the National Forestry and Grassland Science Data Center (NFGSDC). An innovative vibration-assisted machine-vision system was built with alternating back-lit and front-lit illumination, which incorporated a flexible vibratory panel (FVP) to manipulate the multitude of seeds to minimize the occurrence of butting or overlapping, and the lighting from alternating directions to capture phenotypic features both in silhouettes and in vivid color images. To investigate how illumination directions would affect phenotyping, morphological and chromatic metrics were measured, respectively from only the common front-lit images and through the combined use with back-lit images, and applied to distinguish different cultivars and harvest-batches. Results verified that back-lit excelled for reliable segmentation for feature images and accurate morphological metrics, especially when the closeness was clearly revealed in the clustering dendrogram between Nanlin 895 and Zhonglin 46, which shared a common genetic sourcing from P. Euramericana. In contrast, front-lit images were prone to occasional segmentation defects leading to inaccurate morphological measurements due to the highly dynamic range of seed colors, which caused the clustering to lose the genetic relevance. The power of the image-dataset of alternating illuminations was further demonstrated when a decent accuracy of 0.819 yielded from the simple support-vector-machine classification while working on only the back-lit morphological measurements, and the increase to 0.856 with statistical significance if with the addition of chromatic metrics from corresponding front-lit color images, while other image characteristics had been strictly held back. The vibration-assisted alternating illumination protocol established in this work to capture delicate seed-features of Populus cultivars may also be applied to other small grains facing similar imaging challenges, laying a sturdy step-stone of high-throughput phenotyping for large-scale breeding programs and genetic studies.

Keywords: Populus seed, machine vision, camera calibration, flexible vibratory plate, back-lit and front-lit illumination
DOI: 10.25165/j.ijabe.20261901.9850

Citation: Wang X W, Horly M M, Li Z P, Zhao M C, Wu B, Wang M M, et al. High-throughput seed phenotyping of Populus cultivars in China using vibration-assisted machine vision with alternating back-lit and front-lit illuminations. Int J Agric & Biol Eng, 2026; 19(1): 197–212.

1 Introduction

Populus cultivars have long been bred and cultivated in China, and have successfully served many great ecological and national developmental causes, such as forest coverage and providing for the manufacturing sector's overwhelming demand for wood material. To meet the new challenges of improving the efficiency of afforestation and wood production in saline, alkali, or arid regions in China, which combined comprise over one-third of the total land area of China[1,2], a national breeding project has been launched for Populus cultivars with drought, saline, and/or alkali tolerance. Existing Populus cultivars from various parts of China have been

Received date: 2025-04-16 **Accepted date:** 2025-11-10
Biographies: Mubikayi Muhong Horly, MS candidate, research interest: opto-mechatronics, computer vision and automation, Email: mubikayihorly@gmail.com; **Zanpeng Li**, PhD candidate, research interest: opto-mechatronics and deep learning, Email: 3029546474@qq.com; **Maocheng Zhao**, PhD, Professor, research interest: image processing, mechatronics and intelligent systems, Email: mczhao@njfu.edu.cn; **Bin Wu**, PhD, Associate Professor, research interest: digital design and intelligent manufacturing systems, Email: wubin@njfu.edu.cn; **Mengmeng Wang**, MS candidate, research interest: digital design and engineering software simulation, Email: 3310493769@qq.com; **Jiale Deng**, MS candidate, research interest: opto-mechatronics and computer vision, Email: 1195341225@qq.com; **Mengmeng Qiao**, PhD, Lecturer, research interest: spectroscopy and hyperspectral imaging technology, Email: qiaomengmeng2024@njfu.edu.cn; **Xiao Chen**, MS candidate, research interest: discrete element analysis, Email: 1738045174@qq.com; **Yongjie Lu**, MS candidate, research interest: spectroscopic analysis, Email: 2381383841@qq.com; **Liang Qi**, PhD, Associate Professor, research interest: artificial intelligence and computer vision, Email: liangqi@njfu.edu.cn; **Weijun Xie**, PhD, Lecturer, research interest: computer vision, spectroscopic analysis and deep learning, Email: weijun_xie@njfu.edu.cn; **Hongyuan Zou**, Associate Professor, research interest: automation and non-destructive inspection, Email：zouhy@njfu.edu.cn; **Yan Zhang**, Senior Engineer, research interest: physiological cultivation of poplar (Populus spp.), Email: zhangyan913@yeah.net; **Rusheng Peng**, PhD, Professor-level Senior Engineer, research interest: selection, breeding and promotion of Populus (poplar) cultivars, Email: pengrusheng@126.com.
***Corresponding author: Xiwei Wang**, PhD, Lecturer, research interest: spectral imaging, detection technology, and opto-mechatronics of intelligent agricultural equipment. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Longpan Road 159, Xuanwu District, Nanjing 210037, China. Tel: +86-13813881595, Email: wangxiwei@njfu.edu.cn.

collected and new cultivars are being created through both conventional methods and genetic engineering. Meanwhile, poplar phenotyping of various organs, growth stages, and scales are underway to facilitate the quest for these cultivars. This work aimed to conduct a high-throughput image-based phenotyping study of Populus seeds of common cultivars cultivated in China to gain deeper insights into their genetic and physiological traits. P.R. of China has a rich diversity of Populus cultivars, including native, clone, and introduced varieties, such as Zhonglin 46 (*P. x euramericana CL.,* Zhonglin 46), Nanlin 895 (*P. deltoides* × *P. euramericana cv.,* Nanlin 895), and Populus simonii[3]. In recent years, seed phenotyping analysis has primarily relied on three key methodologies. The first approach, manual inspection, is widely used in small-scale studies but is inherently labor-intensive, time-consuming, and susceptible to human error. The second method, biochemical analysis, provides detailed compositional insights but is resource-intensive, time-consuming, and often destructive, making it less suitable for large-scale studies[4,5]. Due to the limitations of these two traditional methods, many scholars have turned to image-based phenotyping. This non-destructive approach supports high-throughput analysis, enabling real-time assessment and capturing multiple seed phenotypic traits at once, making it ideal for seed evaluation. This is the case of Beak et al.[6], who employed image analysis techniques for comprehensive soybean seed characterization. Meanwhile, Van de Looverbosch et al.[7] implemented X-ray CT and deep learning technology to conduct a high-throughput nondestructive sugar beet fruit analysis. Further advancing the field, Huang et al.[8] developed an innovative approach utilizing handheld 3D laser scanning technology for efficient high-throughput phenotyping of legume seeds.

Nevertheless, applying machine vision-based phenotyping technology to the small-sized Populus seeds imposes new challenges, including efficient handling of multitudes of seed for spatial separation from each other to avoid butting or overlapping, and precision imaging to acquire the phenotypic traits very often measured in micrometers.

Some of the machine vision-based phenotypic works that performed successfully on various seeds may be categorized into three groups in terms of the direction of illumination and the nature of image features to be used, as listed in Table 1. However, to the knowledge of the authors, it has not yet been reported how the choice of directions of illumination would affect phenotypic measurements and seed classifications.

Table 1 Three approaches to seed phenotyping in terms of illumination and features

Illumination	Feature	Purpose	Code	Reference
Back-lit	Morphological	Measurement	BL-M	[9,10]
Front-lit	Morphological/textural	Measurement/ classification	FL-M	[11]
Front-lit	Morphological/textural + chromatic	Measurement/ classification	FL-MC	[5,12-16]
Alternating back- and front-lit	Morphological/textural + chromatic	Measurement/ classification	CL-MC	-

Therefore, to best serve the broadest potential phenotyping works, we decided that the phenotypic traits of China Populus seeds should be captured in images from both illuminations when building the seed image-data bank of Populus cultivars across China with China National Forestry and Grassland Science Data Center (NFGSDC)[17,18].

A mechanical vibration-assisted machine vision system (MVS) was constructed for this purpose and used to capture the seeds of Populus cultivar across China in both back-lit silhouettes and front-lit color images. The phenotypic image pairs collected in this work from contrast illuminations provided a precious opportunity to conduct a comparative study of the image-based phenotyping approaches that have been widely adopted in seed research. The following three approaches, therefore, are designated as the baselines: BL-M, standing for back-lit illumination for morphological features; FL-M, front-lit illumination for morphological features; and FL-MC, front-lit for both morphological and chromatic features. More importantly, a fourth approach that combines the merits of back-lit and front-lit lightings for morphological and chromatic features, i.e. CL-MC, was investigated and compared with the baselines.

2 Material and methods

Figure 1 illustrates the workflow, which consists of four stages: stage 1 (a-f) seed preprocessing; stage 2 (g-h) hardware and software system integration; stage 3 (i-j) data acquisition and publishing; and stage 4 (k-l) phenotyping analysis. The seed preprocessing stage includes capsule collection from populus (a-c) stands across China, controlled drying (d), seed-fluff separation (e), and standardized seed storage (f). The stage of hardware and software system integration involves the integration of a novel vibration-assisted machine vision system, camera calibrations (g), and control system development (h). The image acquisition and publishing stage (i-j) involved the acquisition of both silhouettes and color images under respective back-lit and front-lit illuminations, as well as the processing of image-data to meet the quality requirements of the dataset of over 1 187 000 single-seed images at 20 μm/pixel resolution with China National Forestry and Grassland Science Data Center (NFGSDC)[17,18]. Subsequently, in stage 4 phenotyping analysis (k), comparative analysis was conducted on a subset of the databank, in which 11 167 singular seeds of seven cultivars/batches were selected, each of the image pair of a silhouette and a color image. Morphological features were extracted respectively from the back-lit silhouette images (BL-M and CL-M) and the front-lit color ones (FL-M); the chromatic features were measured from the color images but masked, respectively, using the seed segmentation results directly from color images (FL-MC) and those from back-lit images (CL-MC). Then, comparative analysis was performed using clustering dendrograms and support vector machine (SVM)-based classifications between the four approaches to machine-vision based seed phenotyping. This also served the purpose of testing the usability of the Populus seed image-dataset for phenotyping.

2.1 Populus seed collection and preprocessing

Poplar trees (genus Populus) are more widespread in northeastern China[19]. They produce small seeds enclosed in cottony fibers that facilitate wind dispersal. These seeds are released during late spring or early summer and remain viable for only a brief period[20]. The collection, cleaning, and storage of poplar seeds present several challenges. First, harvest begins optimally when a few seed capsules start to split open naturally, indicating peak maturity for collection[21-24]. Collecting seeds requires dealing with the considerable height of poplar trees (Figure 1a)[25]; in this work, a lift truck was used to access the ripened capsules from the tall stands (Figure 1b). Second, to facilitate the release of enclosed seed-fluff from the capsules, all cultivars were exposed to direct sunlight for approximately 5 d. Heat facilitated the natural dehiscence of the capsules, with the degree of opening influenced by their ripeness, thereby enhancing the efficiency of seed extraction (Figure 1c). The

small size of the seeds and their attached cottony fibers posed challenges in handling and processing. To address this, a rapid and efficient removal method was implemented using readily available hardware tools, including brushes and a meshed wire grid. Additionally, the extraction process requires meticulous handling to preserve the delicate seed coat integrity while effectively separating seeds from seed hair (Figures 1d and 1e)[26]. Third, timing is critical for seed collection, as Populus seeds lack dormancy, rapidly losing viability under natural conditions, and may undergo color alterations if harvested either before or after full ripeness. To extend seed viability, proper storage protocols must be implemented, including the use of sterilized materials and appropriate desiccants[27,28]. In this study, seeds from each cultivar/batch were stored in individual glass vials, carefully labeled to prevent mixing (Figure 1f).

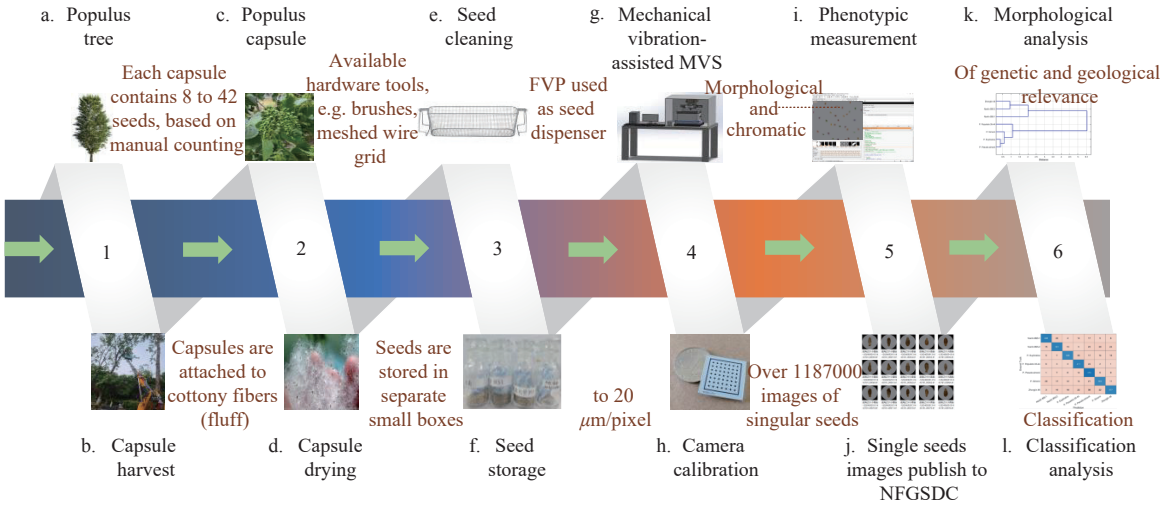


Figure 1 Overall phenotyping process of Populus seed

The genus Populus[3,29-31], which includes numerous species and hybrids, is classified into six sections according to the Eckenwalder system: Abaso, Turanga, Populus (or Leuce), Leucoides, Aigeiros, and Tacamahaca. For a focused analysis, representative cultivars/batches, including three from Aigeiros, one from Turanga, and another three from Tacamahaca, were chosen from a broad collection of poplar hybrids and cultivars across China from Henan, Liaoning, Shanxi, and Jiangsu Provinces, as detailed in Table 2. Notably, Nanlin 895-harvest-1 was collected 9 d earlier than Nanlin 895-harvest-2, a factor that may influence seed characteristics[21,32,33]. Immediately before image acquisition, a small pinch of approximately 60-90 seeds from a sample was manually placed onto the vibratory plate, since precise measuring tools for seeds of this size were unavailable. This procedure was systematically repeated to ensure sufficient image pairs captured for each cultivar/batch, thereby generating a well-balanced dataset for robust statistical analysis. A sample of three singular seeds from each cultivar/batch is illustrated in Figure 4c.

Table 2 Seven selected Populus cultivars across China

Sample	Cultivar release name	Populus section	Origin	Geographical location	Collection time
Zhonglin 46	*Populus* × *euramericana CL., Zhonglin46*	Aigeiros	Introduced	Henan Province	08 May 2024
Populus Euphratica	*Populus euphratica*	Turanga	Native	Liaoning Province	21 May 2024
Populus Pseudo-simonii	*Populus pseudo-simonii*	Tacamahaca	Native	Liaoning Province	22 May 2024
“35-44” Popularis	*Populus* × *popularis, 35-44*	Tacamahaca	Introduced	Shanxi Province	11 May 2024
Populus Simonii	*Populus simonii*	Tacamahaca	Native	Shanxi Province	11 May 2024
Nanlin 895-harvest-1	*Populus deltoides* × *P. euramericanna cv., Nanlin 895*	Aigeiros	Introduced	Jiangsu Province	03 May 2024
Nanlin 895-harvest-2	*Populus deltoides* × *P. euramericanna cv., Nanlin 895*	Aigeiros	Introduced	Jiangsu Province	12 May 2024

2.2 Hardware and software system integration

2.2.1 Hardware system integration

The vibration-assisted dual-lighting image acquisition hardware features the integration of a machine vision system (MVS) with a flexible vibratory plate (Wuxi Danikor Automation Technology Co., Ltd., Wuxi, Jiangsu Province, China). This not only provides the back-lit illumination, which is at 80% of its intensity-capacity in this work, but also lightens the burden of manual efforts to spread the seeds over multiple trials to minimize the occurrence of butting or overlapping, in layouts with uninterrupted and unconnected seed contours, ideal for phenotypic imaging. It includes a high-resolution imagery component, including a wide-angle 12 mm lens for 25-million-pixel imaging (12FA1224-25MP, Phenix Optics, Shangrao, Jiangxi Province, China) working at 2.8 in focal stop and at its minimum working distance of 0.1 m, and a chromatic camera (MV-CS200-10GC, Hikrobot Technology Co., Ltd., Hangzhou, Zhejiang Province, China) working at its full frame resolution of 5472×3648 pixels and 5000 μs in exposure. The front-lit illumination was provided by a ring LED powered by its digital controller (CST-HRS196-W LED, and CST-DPS24120B-4TD controller, Dongguan Kangshida Automation Technology Co., Ltd., Dongguan, Guangdong Province, China), working at 125 of its 256-level intensity capacity.

Back-lit illustration provides for the sharpness and consistency of the contouring of subjects in silhouettes, which is generally regarded as ideal for morphological measurements[34]. Meanwhile, the front-lighting enables the chromatic and textural fidelity as required for the building of the image-data bank of singular Populus

seeds[12].

The U25 flexible vibratory plate (Figure 2) that exempted the manual effort of spreading the multitude of seeds for imaging in proper layouts without butting or overlapping was essentially a POM material panel mounted on four voice coil motors (V_1, V_2, V_3, V_4). The system operates on resonance and wave interference principles, generating coherent vibration fields across the x, y, and z axes. Each motor, Vi, produces oscillations characterized by distinct frequencies ωi, amplitudes Ai, and phases ϕi. Precise control of the system is achieved through independent adjustment of the motors' frequencies (f_i), phase relationships (ϕ_i), and amplitude modulation (A_i), regulated by an internal controller via pulse width modulation (PWM)[35,36]. In practice, vibrational parameters were calibrated using the Populus seeds and configured at: 38-45 Hz in frequency, 40% in power, 4 s in duration for both spreading out and concentrating the seeds, and 0.1 s duration for an interactive shaking with visual feedback to break the butting or overlapping between seeds. Motor phase relationships were maintained at default settings per the manufacturer's recommendation to ensure consistent performance. The FVP was mounted on a steel plate weighing 19.3 kg, which in turn was placed on a bench table. The heavy plate provided sufficient inertia to prevent the FVP from jumping up from workbench in service, meanwhile mitigating system-wide vibration transmitted to the imaging system.

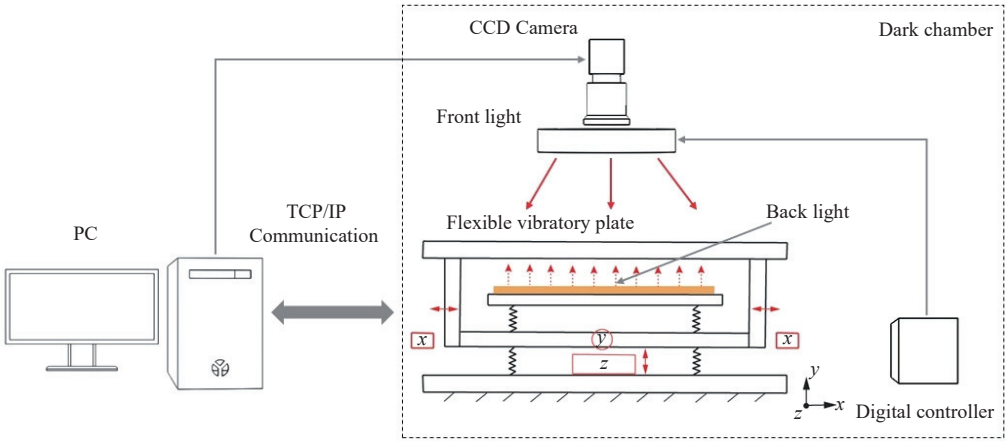


Figure 2 Framework for a vibration-assisted machine vision system

A total of 11 167 seeds, representing a subset from seven cultivars/batches, were processed from an overall pool of 1 187 000 seeds. Each seed was recorded as an image-pair under both back-lit and front-lit illumination conditions to facilitate subsequent measurements and downstream phenotypic analysis. This imaging throughput was supported by a computer equipped with an Intel (R) Core (TM) i9-14900KF processor, 128 GB of RAM, and an NVIDIA GeForce RTX 4080 graphics card, working running the Windows 11 Pro operating system.

2.2.2 Software integration

The integrated software in this work developed for the phenotypic measurement pipeline consists of four core components:

(a) Camera calibration and image acquisition;

(b) Image processing;

(c) Phenotypic measurement; and

(d) Data management, as illustrated in Figure 3.

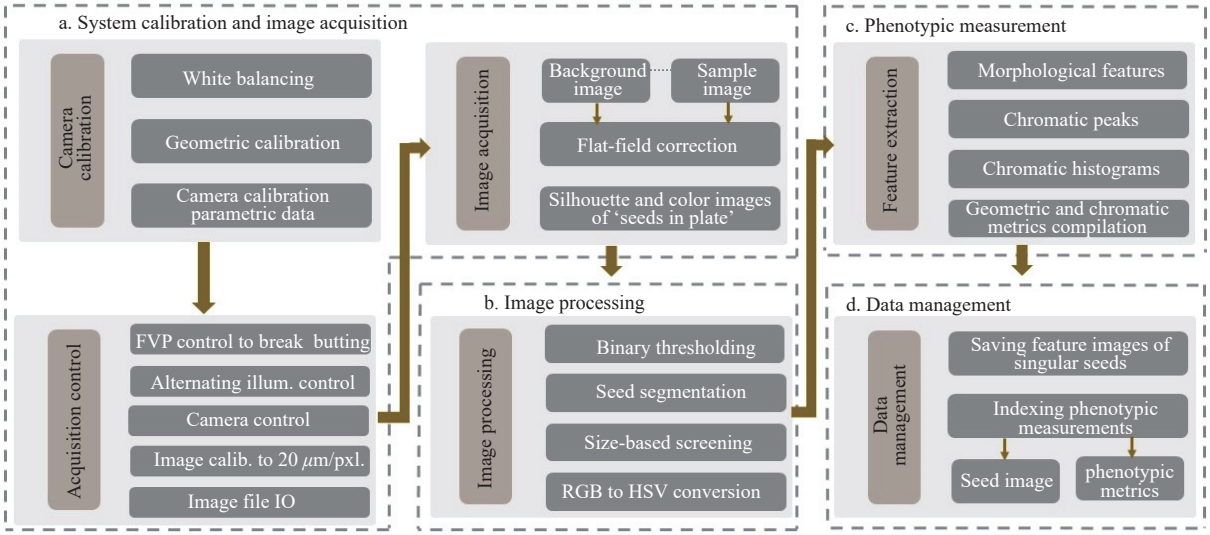


Figure 3 Software integration for feature images and high-throughput phenotypic measurement of Populus seeds

1) Camera calibration and image acquisition

As shown in Figure 3a, the camera was first white-balanced for color fidelity, and then geometrically calibrated for spatial precision using the ceramic standard panel (LIZHI GB025-1-7x7, Dongguan, Guangdong Province, China) and the interactive calibration toolbox from Halcon software (MVTec Software GmbH, Munich,

Germany) for the internal and external camera parameters, with which all captured images were able to be calibrated to exact scale to world coordination in mm.

Essential system integration was developed for acquisition control. This included managing the FVP panel, which positions samples and prevents seed butting (seeds touching each other) or overlaps during imaging. The system also synchronized camera triggering with alternating back-lit and front-lit illumination to improve image clarity. Image calibration was incorporated for automatic correction of geometric distortion and scaling to an exact 20 μm per pixel resolution. File input/output (IO) was also implemented to save images to storage media.

Then, during the image acquisition process, the multitudes of seeds in the FVP plate were able to be manipulated with relative ease via the vibration control to break the butting or overlapping, and the seeds from cultivars and batches were captured with chromatic fidelity, geometric accuracy, and in the exact scale of 20 μm/pixel, in both silhouette and color images of 'seeds in plate'. For the purpose of flat-field correction (FFC), images of blank FVP plate were also saved for both back-lit and front-lit images to be used in the background subtraction to correct the shading and bring the seeds in image corners to the same intensity level of those in the center.

2) Image processing

As illustrated in Figure 3b, the data processing phase begins with the segmentation of seeds from the back-lit or front-lit images after FFC. Specifically, binary thresholding using maximum separability was applied to the gray-valued images of seeds in plate after FFC for the segmentation of seed regions, which were further used for the morphological measurements and as the masks to extract chromatic metrics from corresponding front-lit color images. The binary thresholding for BL-M was applied to the silhouette images from back-lit images, while for FL-M, from the front-lit images. Chromatic values for CL-MC were extracted using the result of the former as mask, and for FL-MC, the latter. This method effectively distinguishes seeds from the background while accommodating variations in seed color.

Another key step in this phase is the rigorous cleaning procedure based on the size and shape-based screening for singular seeds. The algorithm identifies and removes seeds touching borders through image erosion operations, applying a minimum distance threshold of 1 mm from the edges to eliminate partially captured seeds. A size-based screening algorithm is then conducted in two senses, removing objects smaller than 0.5 mm×0.78 mm (typically debris) or larger than 4 mm×1.3 mm (usually poplar seeds butting or overlapping). Using the select shape operation with precise area thresholds scaled to 20 μm/pixel, the algorithm effectively isolates singular seeds.

In addition to these steps, color channel decomposition is applied to separate the image into its individual red, green, and blue (RGB) components, then transformed into the HSV (Hue, Saturation, Value) color space for the following chromatic measurements.

3) Morphological and chromatic measurement

As shown in Figure 3c, the phenotypic measurement was performed for both morphological and chromatic metrics.

The geometric measurement begins with size-related traits, where length (L) and width (W) were computed using minimum bounding and inner circle operations, respectively. Area (A) and contour length (C) provided additional metrics to quantify seed dimensions. Beyond size-related traits, shape descriptors offered additional characterization, such as Aspect ratio (AR) derived as the L/W ratio, and Circularity (I) together with compactness (C_p) assessing from complementing perspectives the deviation from an ideal round shape. Meanwhile, structural metrics, such as anisometry (A_n) and roundness (R_n), quantified seed elongation and symmetry (Table 3).

In addition to geometric traits, chromatic features were also extracted for phenotypic measurement. Histograms for each H, S, and V in 768 values in total were extracted from the corresponding seed-regions to serve as alternative chromatic metrics, providing a chromatic profile of seed pigmentation.

Table 3 Morphological metrics of Populus seeds

Measured morphological characteristics	Description
$A = \Sigma numof pixels$	The area calculated in pixels, providing a measure of seed's size as captured in the digital image[37].
$L = 2\sqrt{(\lambda_1)}$ $W = 2\sqrt{(\lambda_2)}$	Length and width indicate the longest axis and the perpendicular shorter axis, respectively. In the formula, λ_1 and λ_2 are the largest and second largest eigenvalues of the covariance matrix of the seed's pixel coordinates[38].
$C = \Sigma\sqrt{((x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2)}$	Contour length, where (x_i, y_i) are the coordinates of the $i-th$ boundary pixel.
$I = \frac{4\pi A}{perimeter^2}$	Circularity (I) is the measure of how closely the shape of an object resembles a perfect circle[15].
$R_n = \frac{4A}{\pi[Majoraxis]^2}$	Roundness (R) is the measure of how closely the shape of an object approaches that of a mathematically perfect circle[39].
$A_n = Major\_Axis\_Length/Minor\_Axis\_Length$	Its range is ≥ 1, where 1 indicates a perfect circle or square, and larger values indicate more elongated shapes.
$AR = \frac{Width}{Length}$	The aspect ratio provides a basic measure of elongation[40]
$C_p = \frac{C^2}{4\pi A}$	Measures how closely a shape resembles a circle[41]

4) Data management

For data management, as illustrated in Figure 3d, a cropping algorithm created images of singular seeds through cropping from front-lit images using a circular region of interest (ROI) centered at the target seeds that was segmented and screened out in the image processing step, then rotated the seeds to their upright position according to the elliptic axis measurement results. Partials of possible present adjacent seeds were removed through an overpainting with neutral gray values (128, 128, 128).

Since the image segmentation was performed on both the back-lit and front-lit images, and respective results were used for the morphological and chromatic measurement via different phenotyping approaches, feature images of singular seeds and corresponding phenotypic metrics were indexed and stored up specifically for the datasets of BL-M, CL-MC, FL-M, and FL-MC for comparative analysis of the seven representative cultivars/

batches that were selected for the following comparative study on the use of illuminations and morphological/chromatic features.

Only the image-dataset following the CL-MC approach was used for the building of the image-data bank of singular Populus seeds from cultivars/batches, which consisted of over 1 187 000 feature images of singular seeds, securely stored at the NFGSDC, providing free access to the public for Populus seed phenotyping and related studies[18].

2.3 Statistical analysis

2.3.1 Datasets of the four phenotyping approaches per illuminations and features

Both back-lit and front-lit illuminations were intensively used for the phenotyping of seed studies. Back-lit illumination is well-established for obtaining high-contrast silhouettes that enable precise morphological trait extraction[34,42,43], while front-lit illumination remains standard for capturing color and textural information[12,14]. In this work, comparative investigation was done between the two illumination setups and to test the hypothesis that a dual-light configuration—combining back-lighting for morphology and front-lighting for color—would provide a superior feature measurement and better classification accuracy of the seeds from poplar cultivars/batches compared to the use of either illumination alone[44].

Four distinct datasets were generated to compare the combination of back-lit and front-lit illuminations for morphological and chromatic measurements (CL-MC) against the conventional three baselines (as listed in Table 1), as follows.

1) BL-M: Only the morphological features extracted from back-lit images.

2) FL-M: Only the morphological features extracted from front-lit images.

3) FL-MC: Both morphological and chromatic features, extracted from front-lit images.

4) CL-MC: Morphological features extracted from back-lit images, and chromatic features from front-lit images within the segmentation results of seeds from back-lit images.

This design addressed two central questions: (1) Does back-lit illumination enhance the quality of morphological features compared with those extracted using front-lighting (BL-M vs. FL-M)? (2) Does the inclusion of color features from front lighting yield additional gains in classification accuracy under either illumination strategy (FL-MC vs. FL-M; CL-MC vs. BL-M)?

All datasets were subjected to the same statistical analysis pipeline, comprising data standardization and SVM training with optimized hyperparameters, and repeated 30% hold-out cross-validation for performance evaluation. The goal was to observe and compare the differences in classification competence that could be attributed to illumination conditions and the choice of morphological and chromatic feature sets, providing a robust basis for validating the literature-derived premise.

As detailed in section 2.2.2, image processing yielded nine morphological attributes and 768 HSV (Hue, Saturation, Value) color features from both back-lit and front-lit illuminations, the data were organized into the four distinct datasets, i.e. BL-M, FL-M, FL-MC, and CL-MC.

All the statistical data analysis was performed using MATLAB (Version R2024b; The MathWorks, Inc.).

2.3.2 Morphological analysis

1) Clustering

To comprehensively evaluate the seven Populus cultivars, morphological analysis was conducted across three primary trait categories: size-related traits (L, W, A, C), shape descriptors (L/W, I, C_p), and structural metrics (A_n, R_n). For each trait, quantitative metrics—including the minimum, mean (μ), maximum, standard deviation (σ), coefficient of variation (CV, %), and Z-score—were computed. Building on these findings, comparative morphological analysis, clustering analysis, and SVM-based classification were conducted to assess trait variability and classification accuracy. Additionally, a comparative morphological analysis was conducted using the median, interquartile range (IQR), and outlier counts to assess trait distribution across the seven cultivars, as detailed in Table 4. To classify cultivars based on morphological traits and identify the most influential features, Principal Component Analysis (PCA) and hierarchical clustering (HC) were applied. PCA was

Table 4 Summary of median, interquartile range (IQR), and outlier counts across seven Populus cultivars

	Nanlin 895-harvest-1			Nanlin 895-harvest-2			P. Popularis 35-44			P. Pseudo-simonii		
	Median	IQR	Outliers	Median	IQR	Outliers	Median	IQR	Outliers	Median	IQR	Outliers
Length (mm)	2.94	0.481	181	3.22	0.423	134	2.44	0.328	50	2.80	0.251	101
Width (mm)	0.620	0.180	139	0.740	0.180	102	1.12	0.240	25	1.16	0.180	82
Area (mm^2)	1.33	0.443	157	1.83	0.612	107	2.09	0.593	49	2.50	0.469	95
Contour length (mm)	6.72	0.998	185	7.42	0.934	147	6.10	0.795	64	6.92	0.619	101
Compactness	2.58	0.689	205	2.32	0.479	138	1.43	0.163	98	1.51	0.150	133
Circularity	0.171	0.0566	174	0.207	0.0569	91	0.426	0.0841	22	0.377	0.0585	95
Anisometry	4.38	1.43	196	3.97	0.957	141	2.17	0.423	86	2.42	0.368	107
Roundness	1.07	0.100	263	1.05	0.0855	171	1.01	0.0113	112	1.02	0.0148	153
Aspect Ratio	3.80	1.71	220	3.28	1.11	144	1.20	0.442	88	1.47	0.396	109

	Zhonglin 46			P. Euphratica			P. Simonii		
	Median	IQR	Outliers	Median	IQR	Outliers	Median	IQR	Outliers
Length/mm	3.29	0.351	229	2.66	0.260	239	2.75	0.351	208
Width/mm	0.880	0.320	192	1.10	0.160	244	1.00	0.200	195
Area/mm^2	2.21	0.844	196	2.30	0.465	251	2.14	0.496	196
Contour length/mm	7.57	0.868	218	6.55	0.589	250	6.62	0.763	199
Compactness	2.00	0.353	253	1.47	0.143	267	1.60	0.198	226
Circularity	0.247	0.0667	201	0.392	0.0681	231	0.337	0.0710	202
Anisometry	3.47	0.834	259	2.37	0.389	256	2.66	0.511	213
Roundness	1.03	0.0398	282	1.01	0.0106	303	1.01	0.0157	286
Aspect Ratio	2.60	0.934	274	1.39	0.406	263	1.70	0.539	222

used to reduce dimensionality while preserving 95% of the variance, ensuring that the most significant features were retained for clustering. This step mitigated collinearity among traits and improved clustering performance by eliminating redundant information. After dimensionality reduction, Euclidean distances between observations were computed, followed by HC using Ward's linkage method to minimize intra-cluster variance. The optimal number of clusters was determined using the silhouette method, which evaluates the mean silhouette score across different cluster counts.

The morphological metrics that were submitted with the image-data to the NFGSDC were measured via the back-lit illumination, while the measurements were repeated on the front-lit images also for the purpose of comparative analysis for the following sections.

2) Classification

The dataset mentioned in the data availability section was imported from an excel file, then subsequently partitioned into training and testing sets using a stratified 70-30 train–test split to maintain the proportional representation of seed classes. For classification, a multi-class support vector machine (SVM) was implemented with a one-vs-one error-correcting output code framework. After comparative trials with different kernels, the polynomial kernel of order three outperformed the others, as it consistently provided superior discrimination among seed classes and it was employed with a box constraint of 76.61, ensuring effective multi-class classification strategy. Feature standardization was applied to model training to eliminate scale disparities across morphological descriptors. Model performance was evaluated through accuracy, precision, recall, and F1-score, complemented by confusion matrices and feature importance analysis. This combination of a powerful kernel and a precisely tuned constraint was selected to effectively manage the subtle inter-species variances and high intra-species homogeneity characteristic of Populus seed phenotypes.

2.3.3 Classification analysis

Reconstruction Independent Component Analysis (RICA) was used to compress 768 HSV descriptors into 1-8 discriminative chromatic components (CCmps). This unsupervised technique transforms features into a lower-dimensional space where components are statistically independent, effectively filtering noise while preserving essential chromatic information. Then, an exhaustive permutation was done of the four datasets (BL-M, FL-C, FL-MC, and CL-MC), five SVM kernels (Linear, Polynomials of orders from 2 to 4, and RBF), and eight RICA dimensions (1-8 for chromatic components), with 30 independent runs per configuration totaling 2,700 individual SVM models training and evaluating. RICA dimensionality was capped at eight components based on observed performance plateau and subsequent decline, indicating overfitting. From this extensive evaluation, an optimal classifier for each condition of illumination-and-feature group was elected according to the best classification accuracy across all runs, as were the best ones with the addition of more RICA color components (CL-MCs).

This structured approach allowed us to definitively assess whether the illumination using alternating directions improved the classification accuracy for Populus cultivars compared to using either illumination source independently.

2.3.4 Statistical significance analyses

The statistical validation of performance differences between the four datasets (FL-M, FL-MC, CL-M, CL-MC) were validated using a hierarchical non-parametric framework, selected because accuracy scores from repeated cross-validation exhibited non-Gaussian distributions. Following Demšar's (2006) recommendations for classifier comparison with unequal sample sizes[45], the Friedman procedure was applied to accuracy distributions (n = 30 per dataset) to assess statistical significance among median performances. Effect sizes were quantified using Cliff's Delta (δ). Upon detecting significant differences (p<0.05), Nemenyi post-hoc testing was conducted to control family-wise error rates across multiple pairwise comparisons, with results visualized through critical difference diagrams that group non-significantly different conditions (Figure 10).

However, recognizing that the Nemenyi test can be conservative and sensitive to the number of groups, targeted pairwise analyses were subsequently performed using the Wilcoxon signed-rank test. This powerful test was used to directly validate our core research questions: (1) the effect of illumination strategy (FL-M vs. CL-M; FL-MC vs. CL-MC) and (2) the effect of feature set (FL-M vs. FL-MC; CL-M vs. CL-MC). The null hypothesis for each test was that the median difference between paired accuracy distributions was zero. Cliff's Delta was calculated alongside each test to provide a measure of effect size that is independent of sample size. All tests were conducted at a significance level of α=0.05. This multi-tiered statistical design combines a global assessment with both exploratory (Nemenyi) and confirmatory (Wilcoxon) analyses, thereby ensuring robustness, reproducibility, and interpretability of the findings.

3 Results and analysis

3.1 Seed images

Silhouette and color images were collected from respective back-lit and front-lit illuminations, as can be seen from the cropped partial images in Figures 4a and 4b. Over 11 187 000 images of singular seeds, as in the sample of seven cultivars/batches shown in Figure 4c (which were extracted from the front-lit color images but used the segmentation results from back-lit silhouette images as masks, i.e. via the CL-MC approach) were submitted to NFGSDC and are available to the public for free[17,18].

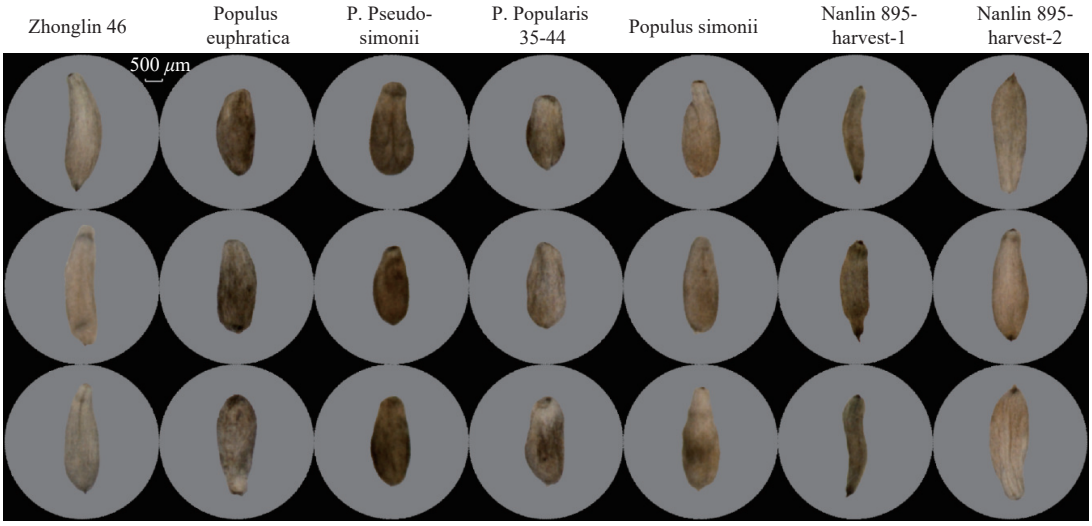
To achieve a balanced sampling of cultivars/batches for reliable analysis, an average of 1500 seeds of the seven populations listed in Table 2 were taken from the NFGSDC Populus seed image-data bank and used for the comparative analysis of the four phenotyping approaches. Figure 5 presents an overview of the seed distribution of the sample dataset.

3.2 Morphological clustering analysis based on back-lit versus front-lit illuminations

As shown in Figure 6a, biologically meaningful clustering patterns were revealed from the morphological clustering dendrogram via the approach using back-lit images (BL-M) where cultivars/batches were grouped also according to geographical origins, indicating correlation between genetic similarity and regional adaptation. This geographical clustering is exemplified by *P. Euphratica* and *P. Pseudo-simonii* from Liaoning Province, which clustered together, demonstrating their shared environmental and genetic influences. Hierarchical clustering showed twelve subclusters within four primary clusters, with the closest association observed between Cluster 4 (Zhonglin 46 and Nanlin 895-harvest-2) and Cluster 6 (Nanlin 895-harvest-1) at a minimal distance of 0.35 (Figure 6a). These cultivars not only shared similar morphological traits including comparable length measurements (3.22 mm vs. 3.29 mm) and aspect ratios (Table 4), but also are all hybrids of *P. Euramericana*, belonging to the same Aigeiros section of populus

a. Silhouette (a crop-out of Zhonglin 46) from back-lit illumination

b. Color image (a crop-out of Zhonglin 46) from front-lit illumination



c. Subset of seven cultivars/batches

Figure 4 Image-data bank of Populus cultivars cultivated across China submitted to NFGSDC

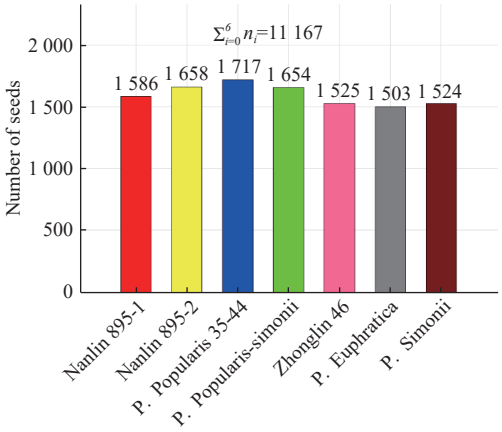
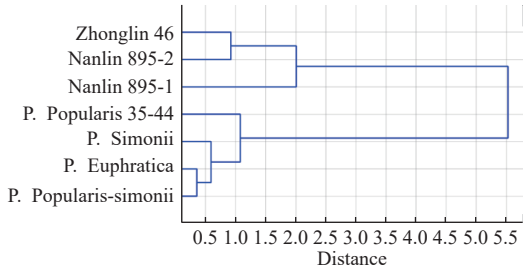


Figure 5 Populus seed distribution

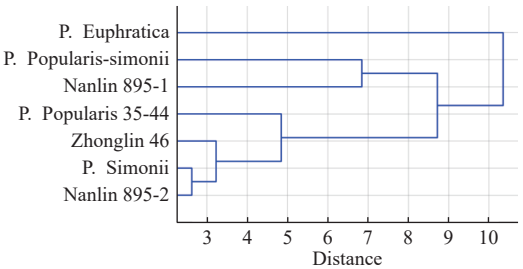
(Table 2), suggesting that clustering was also driven by genetic and phenotypic factors rather than geographical origin alone. The morphological clustering results based on images from back-lit illumination aligned perfectly with literature indications and established taxonomic relationships.

In contrast, morphological dendrogram via the baseline approach using front-lit images (FL-M) presented confusing results having no genetic or geological relevance, despite strong technical performance yielded with four out of seven samples achieving perfect silhouette scores of 1.0. However, the biological interpretation was impossible, as the clustering respected neither geographical origins nor genetic hybridization relationships. The FL-M dendrogram illustrated markedly different linkage patterns, with samples that shared common genetic origins being separated into different clusters in Figure 6b. The stark contrast between illuminations, i.e., FL-M and BL-M, indicates that while computation of clustering functioned smoothly, the FL-M baseline of front-lit approach failed to capture the biological relevance presented in the morphological features of the seeds from the Populus cultivars/batches.

This deviation from prior knowledge of taxonomic relationships suggested some systematic errors may have occurred in the front-lit approach, potentially due to issues with image quality, segmentation algorithms, or preprocessing steps specific to



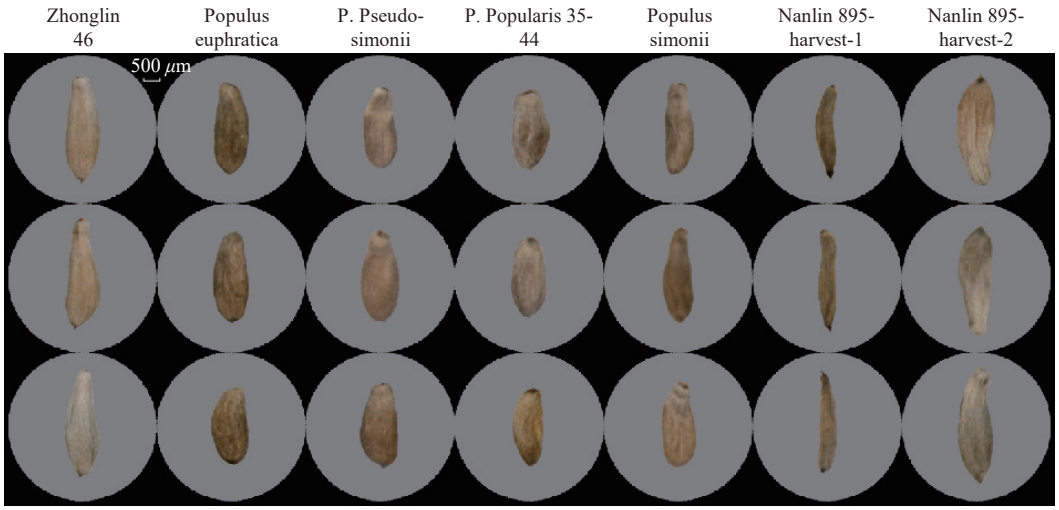
a. From back-lit images



b. From front-lit images

Figure 6 Morphological classification of seven Populus cultivars under different lighting conditions

front-lit images. Tracing the source to the segmentation of singular seeds from respective silhouette and color images, it was found that a small portion of seed masks from the segmentation of front-lit color images had apparent holes, while the ones from back-lit images did not. As shown in Figure 7b, the holes on seed masks in the front-lit approach would corrupt the contours of seeds, in some cases even breaking a singular seed into pieces, consequently poisoning the data quality of morphological metrics of FL-M. Since the chromatic metrics were also extracted within the masked regions from color images, similarly, the color vectors of the same seeds would also differ between the CL-MC and FL-MC approaches. The former utilized the perfect seed masks from FL-M as shown in Figure 7a, but the latter the mistaken masks from FL-M as illustrated in Figure 7b.



a. Perfect segments from back-lit images

b. Segmentation errors from front-lit images

Figure 7 Risk of segmentation error direct from front-lit (FL-M) images, while perfect from back-lit (BL-M) ones

A close comparison of the seeds from Figure 7a and Figure 7b would tell that such holes of uncovered areas occurred mostly on the portion of the highest pixel values, i.e. the brightest parts of a seed. Considering the white-colored material plate of FVP, it is reasonable to infer that it is because Populus seeds are of high dynamic range of brightness and the bright parts of these small fractions of poplar seeds took on an intensity level that was inseparable using the binary thresholding from that of the background under the front-lit illumination. Refer to Figure 4b for the image of material plate in the background of Zhonglin 46 seeds, whose segmentation result is shown in Figure 7b. In contrast, the seed segmentation performed robustly from the back-lit images as shown in Figure 4a. Here, the "bright-field" images guaranteed reliable segmentation, since the consistency in opacity of Populus seeds was far more reliable than brightness, either among the cultivars/batches or on the surface of a single seed.

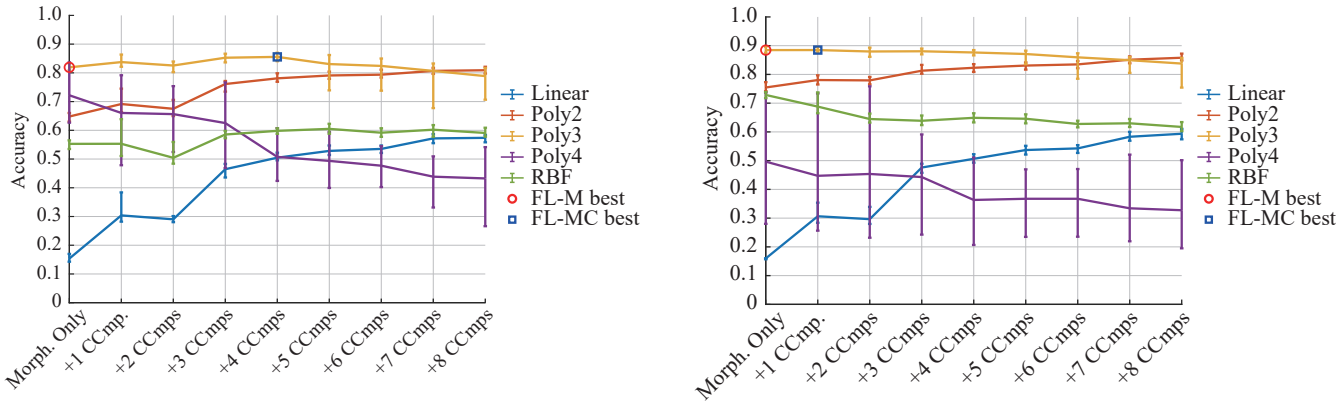
In summary, the reliable segmentation of seeds from back-lit images and morphological metrics derived from it successfully supported the size-and-shape-based analysis to reveal the genetic and geological relationships among the seven cultivars/batches. Therefore, the singular images and measurements submitted to the NFGSDC used the reliable segmentation from back-lit images for seeds and were processed via the CL-MC approach using alternating back-lit and front-lit illumination.

3.3 Classification comparison of illuminations

The classification of cultivars/batches using the reliable seed segmentation from back-lit images for the morphological features and the precision masks to extract the chromatic measurements from front-lit images, i.e., BL-M and CL-M, are shown in Figure 8a, while the ones using only the front-lit images, i.e., FL-M and FL-MC, in Figure 8b. The kernels of linear, polynomials of orders from 2 to 4, and RBF were all investigated, as indicated by the legends. On the *y*-axis to the far left are the accuracies achieved using only the morphological metrics without color information, the ones measured from respective seed segmentations from back-lit images in Figure 8a, and those from front-lit images, in Figure 8b. Since a direct use of the color metrics in 768 HSV values yielded poor accuracy for all datasets, possibly due to the "dimensionality

curse", chromatic values were compressed using the RICA algorithm into color components. Classification results were plotted along the *x*-axis with the incremental number of RICA color components added to the morphological measurements, curbed to the limit of 8 based on the observed performance plateau and subsequent decline, indicating overfitting. Each datapoint showed the mean, max, and min of 30 replicates.

As illustrated clearly in Figure 8a, the choice of classifiers was over feature selection, with kernels of polynomials of order 3 and order 2 performing consistently on the top tier. The former led the classification accuracy in most cases until the adding of the 7th color component, when it was caught up with and surpassed by the latter. From the perspective of phenotypic features, with morphological metrics alone, the 3rd order polynomial kernelled SVM achieved a decent mean accuracy of 0.819, indicating that size and shape alone, i.e., the BL-M approach, could distinguish the seeds from the seven cultivars and batches. With the addition of color information, classification using alternating illumination, i.e., the CL-MC approach, climaxed at 0.856 in mean accuracy. The contribution of color manifested when working with all kernels except RBF by the rise of accuracy curves with the increment of number of color components.



a. From combined use of illuminations (BL-M and CL-MC)

b. From only front-lit illumination (FL-M and FL-MC)

Figure 8 Classification accuracies (in mean, min, and max) using SVM kernels, on measurements of morphological features with the addition of incremental number of color components

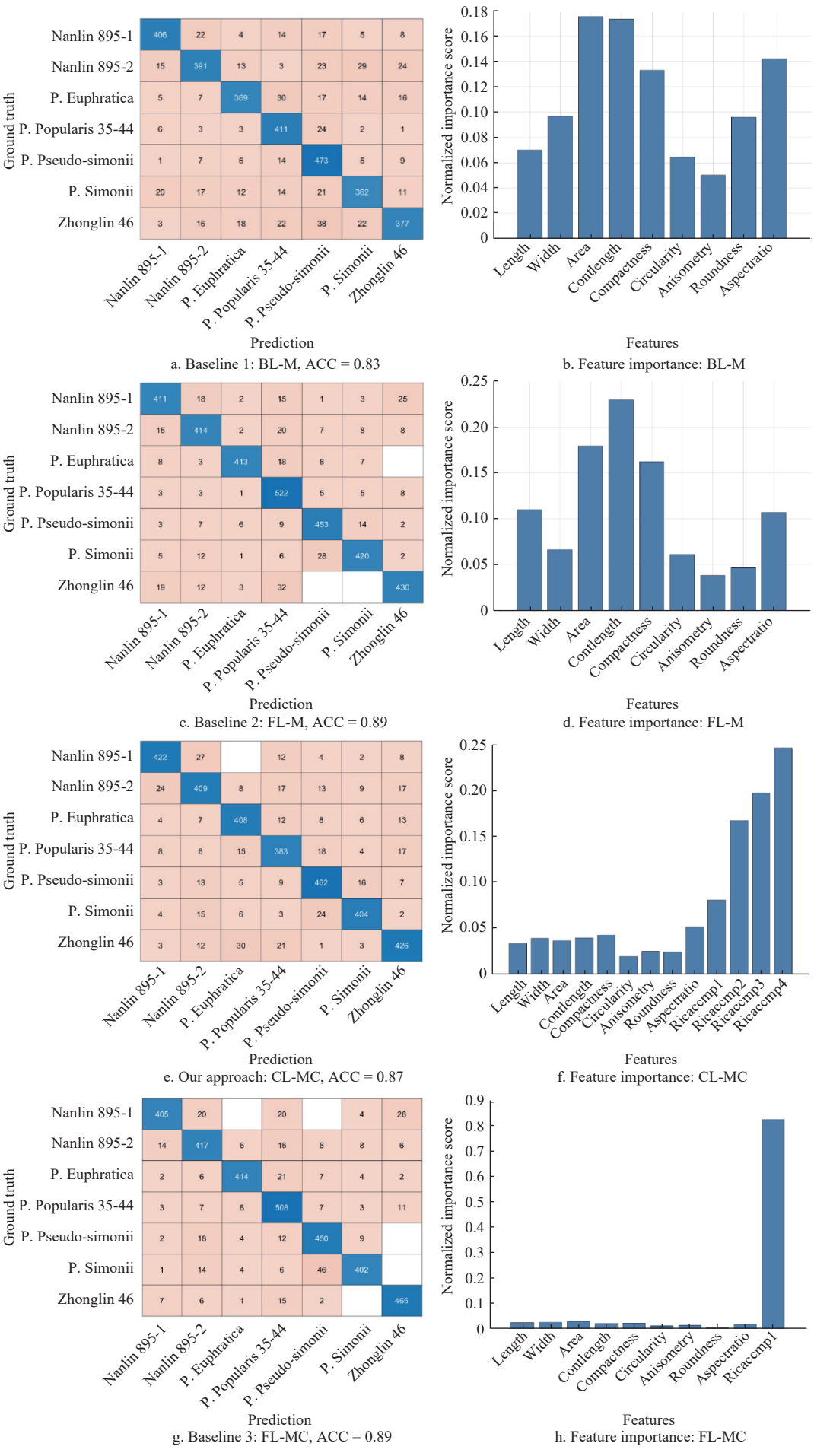
However, the unexpected results in Figure 8b were more intriguing. Though a similar pattern of choice of classifiers over features was also observed, the unreliable segmentation of known errors of possible holes and broken contours identified in section 3.3 and Figure 7b surprisingly boosted the classification performances in almost all cases, except the ones with RBF kernel, over the counterparts using the reliable ones in Figure 7a. It must be that the errors in segmentation had caused more than losing track of the seed contours and making large margins of morphological metrics, but somehow mistakenly bringing in the information that was filtered out during the extraction of size, shape, and color. Maybe the exclusion of highlighted portions of the seeds, as illustrated in Figure 7b, coded the presence and coverage of the zones of extreme brightness into the metrics of area and contour length. Likewise, the chromatic values extracted from a seed would also exclude the highlight zones, making the presence and portion of extreme brightness more prominent in chromatic values.

To verify this, the contribution of features was further investigated in the following sections of best-case comparisons. The model that delivered the best classification accuracy for each phenotyping approach was marked out, with a red disc for the morphological features from back-lit images, BL-M, in Figure 8a, and from front-lit images, FL-M, in Figure 8b. Similarly, the best models making use of both morphological and chromatic features were marked with blue squares for CL-MC, which was the 3rd order polynomial kernelled SVM classifier working on morphological metrics from back-lit with the addition of 4 RICA color components extracted from front-lit images in Figure 8a, and for FL-MC, also a 3rd order polynomial kernel but with the addition of only 1 RICA color component, in Figure 8b. The four champion models were used in the subsequent sections in the best-case comparison of the four phenotyping approaches.

3.4 Best-case comparison of the four phenotyping approaches

Distributions of feature importance of the champion models of the four phenotyping approaches were illustrated in Figure 9, along with corresponding confusion matrices. Comparing the feature-importance distributions of the classification models that considered only morphological metrics, i.e., BL-M and FL-M, as shown in Figures 9b and 9d, the metrics of area and contour length were on the top in contribution to the BL-M approach and they were close to each other. This is understandable because both metrics were size-based and in high correlation. Meanwhile in Figure 9d, the importance of contour length was apparently higher than that of area to FL-M model. The disparity indicates that the broken contours and breakage into pieces resulted from the segmentation errors of front-lit images, which had altered the measurement of contour length more seriously than that of area, due to the mathematical nature that breaking a region or corrupting the contour of the region would cause more change to the sum of perimeters of the new piece(s) than that of area. The prominent bar of contour length in Figure 10d is evidence that the presence and portion of extreme brightness of a seed, as well as the information of textural and brightness natures got through the barrier of size and shape extraction into the morphological metrics of FL-M, leaking extra information that ultimately boosted classification accuracy from 0.83 of the BL-M approach to 0.89.

With the inclusion of chromatic metrics, Figure 9e shows that a proper increase in classification accuracy to 0.87 occurred from the previous 0.83 on only the morphological metrics by BL-M model. Figure 9f shows the contribution made by the gradually increasing importance of color components, which is understandable since the RICA color components were ordered from general to detaile. Combined, the color components outweighed the morphological attributes, indicating the distinct contribution of color to that of size



a. Baseline 1: BL-M, ACC = 0.83
b. Feature importance: BL-M
c. Baseline 2: FL-M, ACC = 0.89
d. Feature importance: FL-M
e. Our approach: CL-MC, ACC = 0.87
f. Feature importance: CL-MC
g. Baseline 3: FL-MC, ACC = 0.89
h. Feature importance: FL-MC

Figure 9 Comparison of phenotyping approaches in classification accuracy and feature importance

and shape in classification. Similarly, the importance of color component outweighed that of morphological metrics combined, as shown in Figure 9h, and there was one more zero square in the confusion matrix of FL-MC indicating one more perfect classification between the cultivars/batches, as shown in Figure 9g. However, the improvement was not discernible in numerical value to the second significant digit. The significance of color contribution will be investigated in the next section per Friedman testing.

3.5 Significance test of differences between the four phenotyping approaches

A Friedman test was conducted to compare the classification accuracies across the four phenotyping approaches, based on the results (Table 5) of 30 runs of the champion model of each phenotyping approach, as detailed in section 2.3.1. The Friedman test showed that there existed a statistically significant difference in performance among the four phenotyping approaches, $\chi^2(3) = 81.040$, $*p* < 0.001$ (Table 6).

Table 5 Friedman test on accuracy distributions

	FL-MC/%	FL-M/%	CL-MC/%	BL-M/%
1	89	90	86	82
2	89	88	87	82
3	88	89	86	82
4	89	89	85	82
5	89	88	86	82
6	88	89	85	82
7	88	89	85	83
8	89	88	86	82
9	88	89	85	82
10	89	88	86	82
11	89	88	85	81
12	89	88	86	81
13	88	89	86	82
14	90	88	86	81
15	88	88	86	82
16	89	88	85	82
17	88	89	85	82
18	88	88	86	82
19	88	89	85	82
20	89	88	85	83
21	89	89	86	82
22	88	88	86	82
23	88	88	84	83
24	88	88	86	82
25	89	88	85	82
26	88	89	85	82
27	89	88	85	82
28	88	89	86	81
29	88	89	86	82
30	88	89	86	81

Table 6 Friedman omnibus test

Test statistic	Degrees of freedom	χ^2	*p*-value	Significance
Friedman test	3	81.0	1.84×10^{-17}	Significant

To further investigate where the differences were, post-hoc Nemenyi analysis and planned pairwise Wilcoxon signed-rank tests were applied, focusing on the effects to classification of illuminations and the incorporation of color features to morphological ones. The results of planned pairwise comparisons, as detailed in Tables 7 and 8, proved positive statistical significance between all pairs, except the one with FL-MC versus FL-M (p=0.885, Cliff's δ=0.485; and Rank difference= 0.667, CD=0.856), verifying that the inclusion of color components made no more improvement in distinguishing the cultivars/ batches via the approach that only used front-lit images.

Table 7 Nemenyi post-hoc pairwise comparisons of differences of average ranks

Comparison	Rank difference	Critical difference	Significant
FL-MC vs. FL-M	0.0667	0.856	FALSE
FL-MC vs. CL-MC	1.47		TRUE
FL-MC vs. BL-M	2.47		TRUE
FL-M vs. CL-MC	1.53		TRUE
FL-M vs. BL-M	2.53		TRUE
CL-MC vs. BL-M	1.00		TRUE

Table 8 Planned pairwise comparisons using Wilcoxon Signed-Rank test

Comparison	Description	*p*-value	Effect Size (Cliff's δ)	Significance
FL-MC vs FL-M	Effect of Adding Color (Front light)	0.885	0.485	Insignificant
CL-MC vs BL-M	Effect of Adding Color (Back light)	1.73E-06	0.00	Significant
FL-M vs BL-M	Effect of Illumination (Morphology only)	1.73E-06	0.00	Significant
FL-MC vs CL-MC	Effect of Illumination (Morphology + Color)	1.73E-06	0.00	Significant

Meanwhile, adding color metrics to the morphological metrics from back-light illumination resulted in a substantial and significant boost in classification performance (CL-MC vs. BL-M: p<0.001, δ=1.000), with the inclusion of color components (CL-MC) outperforming the morphological-only (BL-M) approach in every single run, as listed in Table 5.

As shown in Figure 10, the Nemenyi critical distance was 0.856, meaning that any difference of average ranks must be greater than 0.856 to be of statistical significance. Referring to the score sheet of four champion models in Table 7, the requirement was not met by the pair of FL-MC versus FL-M, and they were marked with the thick line connecting them in Figure 10. On the other hand, the inclusion of front-lit color information provided by the alternating illumination via the CL-MC approach performed consistently better than when using only the morphological metrics from back-lit images via the BL-M approach. The consistency was evidenced by the average ranks of 3.0 and 4.0, two integers, meaning that in the ranking of all sets of comparisons between the two approaches, CL-MC was always better than BL-M, with no exceptions.

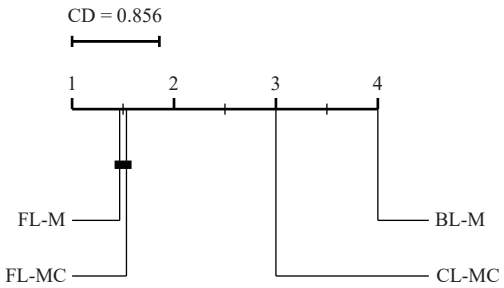


Figure 10 Nemenyi Critical Distance Test of the four phenotyping approaches all showed significant differences except the pair of FL-M versus FL-MC

It cannot be ignored that both front-lit approaches outperformed in classification accuracy. However, the outperformance of the front-lit approaches of FL-M and FL-MC using only the front-lit images over BL-M or CL-MC should not be interpreted as the effect of illuminations, but rather that of an error of segmentation in front-lit images and its propagation into the morphological and chromatic feature extracted for both FL-M and FL-MC, especially when the seed segmentation from front-lit images mistakenly let textural and brightness information through the barrier of morphological and color extraction. Once the extra information got through to work for the classification, the advantage was so overwhelming that the inaccuracy of morphological metrics induced by the occasional occurrence of segmentation errors from front-lit images became negligible.

The non-significance of the difference between FL-M and FL-MC, as Figure 10 shows, further demonstrated that the importance of this extra information leaking through during the front-lit image segmentation was so paramount that the accompaniment of either morphological or chromatic information made no difference to the cultivar/batch classifications.

4 Discussion

4.1 Coma preprocessing for Populus seed

Like many other seeds, an air drying process would facilitate their separation from the encapsulating capsule[46,47], but there is no standard drying and extraction procedure in existence for Populus seeds. Due to the lightweight and fluff-covered nature of Populus seeds, conventional sun-drying methods are unsuitable as seeds are easily dispersed by air currents and sunburn might parch the coma, drying it out before the internal cottony seeds are able to get released. To overcome these challenges, an alternative air-drying method was employed: seed capsules were dried in paper bags under natural sunlight to minimize mechanical damage and preserve physiological integrity, as Sylva did with Cactaceae seeds[48]. Following best practices, drying was completed within ten days of collection[20,49]. Bags were periodically repositioned to ensure uniform drying and prevent moisture retention or overheating.

Capsules began releasing fluff within 2-4 d, as observed by Smreciu et al.[20], with variation depending on temperature, cultivar, and geographic origin. The optimal temperature range for fluff release in our study was 20°C-25°C, slightly higher than the laboratory-controlled 20°C-24°C reported by Bhutta et al.[26], likely due to outdoor peak temperatures reaching 27°C. The minor 2°C discrepancy may have contributed to the loss of some seeds due to excessive heat exposure. Some capsules experienced sunburn, trapping seeds within dried cotton fibers, highlighting the critical importance of temperature control during post-harvest processing.

The period of viable time of Populus seed is known to be from two to four weeks[50], varying by cultivar, season, and geography. For long-term storage, subfreezing temperatures (2°C-5°C) and desiccant-sealed glass vials are recommended to extend viability beyond five years[20,26,51]. Desiccants prevent morphological degradation and maintain seed integrity. The drying and extraction processes must be meticulously conducted not only for seed preservation, but also to mitigate the health risks caused by the airborne fluffs, which can cause eye irritation, skin rashes, and workspace contamination[52]. Operators require personal protective equipment including masks, gloves, and dust covers.

The protocol used in this work reduced processing time, enhanced safety, and prevented contamination, confirming similar practices by other researchers who use customized grid mesh systems with hole sizes suited to different seed dimensions[20] and vacuum-based cleaning methods[22,28,50]. However, the low seed recovery yield remained unchanged. This is because of the labor-intensive manual cleaning required by these methods.

4.2 Vibration-assisted phenotypic imaging using alternating illumination

The vibration assistance to break the locks of seed butting or overlapping was critical when working with the multitude of tiny Populus seeds to have each of them properly imaged with complete and clean contours for reliable phenotypic measurements. The vibration-assisted imaging system marked a significant departure from manual positioning of seeds, such as the manual separation of legume seeds for scanner-based analysis by Huang et al.[8], or the manually operated dispensers used for fruit and seed analysis with X-ray computed tomography by Liu et al.[53] The FVP, controlled via a graphical user interface, not only quickened the image acquisition process but also significantly improved the precision of seed alignment[54,55]. Additionally, precise customization of vibrational motion along the x, y and z axes could be tailored to the seed physical properties, e.g., weight and bounciness, along with the estimated batch size through adjusting the frequencies ωi, amplitudes Ai, and phases ϕi of each coil motor Vi. The frequencies, amplitudes, and phases are the primary parameters influencing the dynamic forces, as demonstrated by Wu et al.[56], in their study on the design and experimentation of a vibration plate-based Camellia fruit picking machine. This adaptability ensured optimal handling of Populus seeds, quickening the phenotyping process with efficiency. From the perspective of benefiters, if not for the vibration-assistance, the image-acquisition process for the hundreds of thousands of images of singular Populus seeds in both back-lit silhouette and front-lit color images would have been many hours longer.

Populus seeds displayed color and brightness in a high dynamic range, both across and within the cultivars/batches, similar to the seed texture observed in apples and apricots, as described respectively by Ropelewska et al.[57,58] This variation can lead to contrast and saturation issues when placed against either a white or dark background. To address this challenge, we turned to the application of an alternating back-lit and front-lit illumination. The key motivation for using front lighting is that back lighting lacks the color information necessary to color-based analysis. Morphological measurements could be, and have been, reported in previous seed research performed under front lighting. However, risks of segmentation errors would make the phenotypic metrics gaged this way unreliable when working with seeds with inconsistent contrast to the background in brightness, saturation, or color, as manifested in this work with Populus seeds and demonstrated by the FL-M and FL-MC approaches to phenotyping. Therefore, we built the system to provide both lighting modes: front lighting for color and back lighting for reliable morphological assessment on bright-filed silhouette images[9,10,34]. It is also noteworthy that similar capacities of both back-lit and front-lit illumination had been previously reported for seed phenotyping works, but for different purposes[42,44]. These works either used the back-lit as a supplementary light source to the front-lit illumination to be used at the same exposures, or merely treated the front-lit and back-lit images as stand-alone sources rather than the use of alternating illuminations for the combined morphological and color information, similar to the fourth phenotyping approach of CL-MC that has been explored in this work.

This alternating lighting approach proved to be a crucial aspect

of Populus seed phenotyping that delivered accurate seed image segments and reliable morphological and chromatic feature-metrics, which enabled the successful classification between the seven cultivars/batches with decent accuracy using simple SVM-based modeling.

The alternating back-lit and front-lit illumination approach applies to the phenotyping of other seeds, except where the seeds should be of high transparency, in which case back lighting alone could possibly suffice.

4.3 Forestry interpretations

During the analysis phase, two phenomena were observed.

First, unlike the Liaoning cultivars, which were clustered closely based on their geographical provenance, aligning with findings by Medina et al.[59] that the 25 varieties of quinoa seeds from the same place clustered together, the Nanlin cultivars were not directly clustered together within a single group. Nanlin 895-harvest-2 from Jiangsu Province was found to directly associate with Zhonglin 46 from Henan Province, exhibiting the closest inter-cluster distance, rather than with Nanlin 895-harvest-1 as initially expected. This result suggests that the inter-cluster distance is not solely influenced by geographic factors[60,61]. A potential factor contributing to this outcome is the hybridization effect, which results in shared characteristic traits between the two hybrid cultivars. Both Nanlin 895-harvest-2 and Zhonglin 46 are hybrids of Populus x euramericana, sharing similar morphological traits such as elongation and compactness.

This first hypothesis observation coincides with the studies of Deok Han et al.[62] and Juma et al.[63] , which respectively emphasized the significance of morphological traits-based selection rather than geographical environmental influences, and the characterization of Tanzanian avocados influenced by morphology because of the absence of distinct clustering of samples based on their regions of origin.

The second hypothesized cause is the harvest time. Differences in seed maturity and collection time influenced seed color and morphological traits. Nanlin 895-harvest-1 seeds, harvested earlier, were tender, soft, and lighter in color, with shape descriptor metrics surpassing those of Nanlin 895-harvest-2, increasing their inter-cluster distance. A nine-day difference between batches of Nanjing 895 further highlighted this effect, as earlier seeds had a noticeably lighter color and larger width. These variations reflect the natural maturation process of poplar seeds, where lignin accumulation in the seed coat increases while moisture content decreases as seeds ripen. The variations in seed color and morphology corresponding to different maturation stages and collection times align with the observations reported by D. Han et al.[64] in maize: Seeds harvested earlier, such as those from early-maturity maize varieties, attain their ideal harvest time (IHT) sooner than late-maturity varieties while maintaining similar moisture content levels at IHT. This indicates that early-harvested seeds may exhibit distinct physical traits due to their shorter maturation process. The same observation has been made by Del Porte Morales et al.[65] in coffee cultivation, where seeds harvested later in January exhibited superior viability and physical quality compared to those collected earlier in December. Similarly, with peas, seeds harvested later were heavier, had higher protein content, and exhibited improved germination rates[65]. However, more work needs to be done to test the hypothesis of impact of different harvesting periods across at least three dimensions: 1) within the same variety and region, 2) across the same variety but different regions, and 3) across different varieties and regions.

4.4 Phenotypic potential

It was expected that the reliability of seed segmentation from back-lit silhouette images, evidenced by the consistently perfect image segments from the back-lit images, would improve the classification accuracy of cultivars/batches due to the reliably accurate morphological metrics. However, results showed otherwise.

As was pointed out in section 3.5, the perfect image segmentation from back-lit silhouettes and the consequent measurements of phenotypic features (strictly of only the morphological and chromatic ones) did also a good job of keeping other phenotypic features, e.g. textural characteristics, away from the classifiers. The occasional segmentation errors in front-lit images on the bright seed partials might have mistakenly brought in this extra information encoded into the morphological metrics measured through some unclear transfer function therefrom via the front-lit approaches of FL-M and FL-MC, gaining an information advantage over the back-lit BL-M and the combined approach of CL-MC where the unintended features were perfectly filtered out.

However, the decent classification accuracy achieved using only the morphological metrics from back-lit images (BL-M) and the increase with the addition from the chromatic metrics from combined use with the front-lit images (CL-MC) had already proved the potential phenotypic values of the seed image-dataset of alternating illumination, while phenotypic features of other image characteristics had been strictly kept away from the conventional machine learning classification algorithm. When other image characteristics leaked through, classification performance was boosted. It is reasonable to infer that the true potential of this image-dataset would be tested and unleashed using modern classification algorithms based on deep-learnings that harness every bit of its graphical and magnitude capacity.

4.5 Practical implementations

The vibration-assisted high-throughput phenotyping using alternating back-lit and front-lit illumination of small seeds takes on a systematic perspective that considers not only the measurement from images, but also the way morphological and chromatic traits of small seeds can be captured in images with both sharper contours and higher chromatic fidelity. Still further, it considers how the manipulation of seeds in panel into non-contact and non-overlapping layouts could be facilitated to save the time and effort for image acquisition. When the number of seeds reaches the hundreds of thousands, relying solely on manual operations would be much more difficult, if not impossible.

The strength of our vibration-assisted dual lighting machine vision system may be best used for the phenotypic measurement of various small-sized seeds of food, trees, or grass that require spatial precision at tens of micro-meters and high in quantity involving a multitude in the hundreds of thousands, as in large-scale genetic studies and breeding projects.

4.6 Limitations and future work

Although promising efficiency was shown in this work of the vibration-assisted machine vision system in shaking the multitude of seeds in non-contact and non-overlap layouts for imaging and precision phenotypic measurements of Populus seeds using alternating illumination, several limitations should be considered to make future steps toward the goal of a fully automatic phenotyping of small-sized seeds.

First, for extremely small seeds, such as populus seeds, removing all grains from the panel at the end of measurement can be tricky. For example, moving a particle as light as 2.3×10^{-3} g and

as large as 0.6 mm×0.4 mm×1.5 mm, the average size and weight of a single populus seed, air viscosity becomes non-negligible. In particular, during a side-way clearing process near or even on the discharge side of the FVP, where the pair of voice coil motors underneath is operating at very low, if not, some seeds may remain idle or lag. This could likely cause incomplete discharge of seeds between subsequent batches and mixing of specimens.. A possible solution might to equip the discharge side with a passageway that provides vacuum airflow to carry away straying seeds, securing reliable panel clearance between measurements and making a continuous phenotyping process complete, including both seed feeding and, more importantly, reliable discharging of seeds from the FVP plate.

Second, the vibrational configuration of the FVP needs to be tuned before the measurement of a new specimen, if its resonance properties significantly vary from the previous one. Since the driving parameters need to be adjusted properly for all four voice coil motors for the base driving movements, i.e. horizontal, vertical, or centered aggregating and dispensing, and translations toward the four directions, a manual operation by an inexperienced hand could sometimes get prolonged. One possible solution lies in the latest development of artificial intelligence. An automatic tuning module that harnesses the power of spatial-and-temporal deep learning technology monitors and judges from the response of target seeds in plate, which is captured in consecutive image-frames during the sweeping of vibrational configurations, e.g., exciting frequencies and actuating powers, so that the driving parameters of the voice coil motors could possibly be tuned with optimization algorithms both per target specimens and per driving movements. This way, the re-configuration of FVP necessary before the measurement of new specimens of varying resonance properties could be done more consistently and thoroughly with automation and no longer depend on the experience and patience of a human operator.

5 Conclusions

The importance of vibrational assistance in handling the multitude of seeds for quality imaging with minimal occurrence of butting or overlapping cannot be overemphasized. Otherwise, the image acquisition for the over 1 187 000 feature images of singular seeds of Populus cultivars across China to build the image-data bank with NFGSDC, National Science & Technology Infrastructure of China, would have taken many more hours.

The consistency of background-contrast provided by back-lit illumination against the highly dynamic range of seed colors was ideal for robust image segmentation of the seeds of Populus cultivars/batches, for reliable morphological measurements, and for the masks to accurately extract color information from front-lit images, so that alternating illumination was able to deliver the reliability and accuracy demanded by the quality imaging for seed phenotyping.

The potential phenotypic values of the seed image-dataset of alternating illumination were evidenced by the close clustering of the hybrid cultivars sharing a common genetic source using only the morphological metrics reliably measured from back-lit images. The measurement had a decent 0.819 average accuracy based on the same metrics to distinguish the seven cultivars/batches yielded from a simple SVM classification. Measurement accuracy further increased to 0.856 with statistical significance when combining the chromatic metrics extracted from corresponding front-lit images while other image features were strictly held back.

The vibration-assisted machine-vision system with alternating back-and front-lit illumination in this work established an efficient and reliable protocol to capture the delicate phenotypic features of the seeds of major Populus cultivars in China. It may also be applied to the phenotyping of other small grains facing similar challenges of multitude and illumination, thus laying a sturdy step-stone of high-throughput phenotypic measurement for large-scale breeding programs and genetic studies.

Acknowledgements

This work was supported by Biological Breeding-National Science and Technology Major Project (Grant No. 2023ZD0405605-1). The authors gratefully acknowledge this financial support, which enabled the development of the vibration-assisted machine vision system, the collection and processing of poplar capsules from eastern, northeastern, and central China, and the execution of high-throughput phenotyping analysis central to this study. The authors also extend their sincere thanks to Prof. Fu Fangfang for the seed samples used in this work and Mr. Darko George for his consultation on geographic information processing.

[References]

[1] Yun X, Chen Y. International development of saline-alkali land and its enlightenment to China. Territory & Natural Resources Study, 2020; 1: 84–87.

[2] Zhang Q, Huang J, Yang J, Guan X, Yu H, Zhu B, et al. Advances in research on climate change and its effects on the arid and semi-arid regions of China over the past century. J Meteorol Res, 2025; 39(3): 673–687.

[3] Yang Z L, Zhou S Y, Zhang W D, Yang Z X. Poplar genetic resources in North China: the challenge of sustainable forestry. Forest Genetic Resources (FAO), 1999; No. 27. https://www.fao.org/4/x4133e/X4133E02.htm

[4] Liu S, Huang Z, Xu Z, Zhao F, Xiong D, Peng S, et al. High-throughput measurement method for rice seedling based on improved UNet model. Computers and Electronics in Agriculture, 2024; 219: 108770.

[5] Tu K, Wu W, Cheng Y, Zhang H, Xu Y, Dong X, et al. AIseed: An automated image analysis software for high-throughput phenotyping and quality non-destructive testing of individual plant seeds. Computers and Electronics in Agriculture, 2023; 207: 107740.

[6] Baek J, Lee E, Kim N, Kim S L, Choi I, Ji H, et al. High throughput phenotyping for various traits on soybean seeds using image analysis. Sensors, 2020; 20(1): 248.

[7] Van De Looverbosch T, Vandenbussche B, Verboven P, Nicolaï B. Nondestructive high-throughput sugar beet fruit analysis using X-ray CT and deep learning. Computers and Electronics in Agriculture, 2022; 200: 107228.

[8] Huang X, Zheng S, Zhu N. High-throughput legume seed phenotyping using a handheld 3D laser scanner. Remote Sensing, 2022; 14(2): 431.

[9] Lin T Y, Huang P C, Syu C Y, Huang K Y. Automatic inspection system for red paddy seeds. Sensors and Materials, 2023; 35(12): 4383–4395.

[10] Ozkaya Y A. Digital image processing and illumination techniques for yarn characterization. J Electron Imaging, 2005; 14(2): 023001.

[11] Zu Q, Liu T, Zhu W, Pan Y, Wang J, Song X, et al. Automated seed counting using image processing and deep learning. Front Plant Sci., 2025; 16: 1659781.

[12] Loddo A, Loddo M, Di Ruberto C. A novel deep learning based approach for seed image classification and retrieval. Computers and Electronics in Agriculture, 2021; 187: 106269.

[13] Mishra P, Lohumi S, Ahmad Khan H, Nordon A. Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches. Computers and Electronics in Agriculture, 2020; 178: 105780.

[14] Chopin J, Kumar P, Miklavcic S J. Land-based crop phenotyping by image analysis: consistent canopy characterization from inconsistent field illumination. Plant Methods, 2018; 14(1): 39.

[15] Dayrell R L C, Ott T, Horrocks T, Poschlod P. Automated extraction of seed morphological traits from images. Methods Ecol Evol, 2023; 14(7): 1708–18.

[16] Halcro K, McNabb K, Lockinger A, Socquet-Juglard D, Bett K E, Noble S D. The BELT and phenoSEED platforms: shape and colour phenotyping of seed samples. Plant Methods, 2020; 16(1): 49.

[17] National Science & Technology Infrastructure of China, National Forestry and Grassland Science Data Center (NFGSDC). Singular seed visible image dataset of poplar species from eastern, northeastern, northern, and central China. 2025; Available: https://www.forestdata.cn/dataDetail.html?id=aed290f4-60dd-4812-bfe6-53d396559846
[18] National Science & Technology Infrastructure of China, National Forestry and Grassland Science Data Center (NFGSDC). Seeds on tray under back-light and front-light illumination visible image dataset of poplar species from eastern, northeastern, northern, and central China. 2025; Available: https://www.forestdata.cn/dataDetail.html?id=01946496-999a-4883-89d1-b7db3335b84e
[19] Liu X, Hou L, Ding C, Su X, Zhang W, Pang Z, et al. Effects of stand age and soil microbial communities on soil respiration throughout the growth cycle of poplar plantations in northeastern China. Front Microbiol, 2024; 15: 1477571.
[20] Smreciu A, Landhäusser S, Marenholtz E, Sobze J M, Gould K, Schoonmaker A. Aspen seed collection and cleaning. 2013; Available: https://www.cclmportal.ca/sites/default/files/2020-02/Aspen_Seed_Collection_and_Cleaning.pdf
[21] Brown K R. Catkin growth, seed production, and development of seed germinability in quaking aspen in central Alberta. Tree Planters' Notes, 1989; 40(2): 25–29. https://digitalcommons.usu.edu/aspen_bib3209
[22] Fung M Y P, Hamel B A. Aspen seed collection and extraction. Tree Planters' Notes, 1993; 44(3): 98–100.
[23] Johnson L. Effect of humidity on the longevity of Populus and Ulmus seeds in storage. Canadian Journal of Research, 1946; 24(6): 298–302.
[24] Maisenhelder L C. Planting and growing cottonwood on bottomlands. State College, Miss.: Mississipp.Agricultural Experiment Station, 1951. Available: https://catalog.hathitrust.org/Record/011461492
[25] Cui B X, Liu J L, Zhou N, Wu W P, Tang X H, Ding C J, et al. Effect of thinning intensity on fiber morphology and crystallinity of poplar. BioRes 2024; 20(1): 1024–1036.
[26] Bhutta N, Nunez-Martinez O F, Mei C, Bräutigam K. Seed collection in temperate trees—clean, fast, and effective extraction of populus seeds for laboratory use and long-term storage. Bio-protocol, 2024; 14(3): e4927.
[27] Einspahr D, Schlafke D. A method for Aspen and Cottonwood seed extraction. Tree Planters' Notes, 1957; 28: 10.
[28] Roe E I, McCain D P. A quick method of collecting and cleaning Aspen seed. Tree Planters' Notes, 1961; 51: 17–18.
[29] Liu Y, Wu X, Zhang J, Liu S, Semple K, Dai C. Maturation stress and sood properties of poplar (Populus × euramericana cv. 'Zhonglin46') tension wood. Forests, 2023; 14(7): 1505.
[30] Weisgerber H, Han Y. Diversity and breeding potential of poplar species in China. The Forestry Chronicle, 2001; 77(2): 227–237.
[31] Borkhert E V, Pushkova E N, Nasimovich Y A, Kostina M V, Vasilieva N V, Murataev R A, et al. Sex-determining region complements traditionally used in phylogenetic studies nuclear and chloroplast sequences in investigation of Aigeiros Duby and Tacamahaca Spach poplars (genus *Populus* L. , Salicaceae). Front Plant Sci, 2023; 14: 1204899.
[32] Faust M E. Germination of Populus grandidentata and P. tremuloides, with particular reference to oxygen consumption. Botanical Gazette, 1936; 97(4): 808–821. Available: http://www.jstor.org/stable/2471529
[33] Junaidi A, Andriyanto M. Seed collection time effect on the germination rate and growth of rubber tree rootstock. In: Proceedings of the 3rd KOBI Congress, International and National Conferences (KOBICINC 2020). Atlantis Press; 2021; pp.278–282. doi: 10.2991/absr.k.210621.046
[34] Zafari S, Eerola T, Sampo J, Kalviainen H, Haario H. Segmentation of overlapping elliptical objects in silhouette images. IEEE Trans on Image Process, 2015; 24(12): 5942–5952.
[35] Han L, Deng J. A study on flexible vibratory feeding system based on HALCON machine vision software. Zhengzhou, China, 2015. Available: https://www.atlantis-press.com/article/18191 Accessed on [2024-09-16]
[36] Rao S S. Vibration of continuous systems. 1st ed. Wiley; 2019. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119424284Accessed on [2025-11-20]
[37] Steinecke C, Lee J, Friedman J. A standardized and efficient technique to estimate seed traits in plants with numerous small propagules. App. Plant Sci., 2023; 11(5): e11552.
[38] Ghimire A, Kim S H, Cho A, Jang N, Ahn S, Islam M S, et al. Automatic evaluation of soybean seed traits using RGB image data and a python algorithm. Plants, 2023; 12(17): 3078.
[39] Dayrell R L C, Ott T, Horrocks T, Poschlod P. Automated extraction of seed morphological traits from images. Methods Ecol Evol, 2023; 14(7): 1708–1718.
[40] Cervantes E, Martín J J, Saadaoui E. Updated methods for seed shape analysis. Scientifica, 2016; 2016: 1–10.
[41] Florczyk S. Video based indoor exploration with autonomous and mobile robots. Journal of Intelligent and Robotic Systems, 2005; 41(4): 245–262.
[42] Baek J, Lee E, Kim N, Kim S L, Choi I, Ji H, et al. High throughput phenotyping for various traits on soybean seeds using image analysis. Sensors, 2020; 20(1): 248.
[43] Tu K, Wu W, Cheng Y, Zhang H, Xu Y, Dong X, et al. AIseed: An automated image analysis software for high-throughput phenotyping and quality non-destructive testing of individual plant seeds. Computers and Electronics in Agriculture, 2023; 207: 107740.
[44] Hernández S, Zhong V, Brophy J A N. SeedSeg: image-based transgenic seed counting for segregation analysis of T-DNA loci. Plant Methods, 2025; 21(1): 87.
[45] Demšar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research, 2006; 7(1): 1–30. Available: https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf
[46] Barrozo M A S, Mujumdar A, Freire J T. Air-drying of seeds: A review. Drying Technology, 2014; 32(10): 1127–1141.
[47] Corbineau F. The effects of storage conditions on seed deterioration and ageing: How to improve seed longevity. Seeds, 2024; 3(1): 56–75.
[48] Silva J H C S, Azerêdo G A D, Souza V C D. Conservation of seeds of cactaceae species endemic to the caatinga biome: Pilosocereus pachycladus and Tacinga inamoena. Rev Caatinga, 2023; 36(1): 115–123.
[49] Tiebel K, Dahlmann J, Karge A. Global warming could shorten the seed lifespan of pioneer tree species and thus natural regeneration window of damaged areas. European Journal of Forest Research, 2024; 143(2): 437–450.
[50] Wyckoff G, Zasada J C. Populus L, Bonner F. Woody plant seed manual. Agric Handbook, 2005; 727: 856–871. Available: https://www.fs.usda.gov/nsl/Wpsm/Populus.pdf
[51] Young J A, Young C G. Seeds of woody plants in North America. Rev. and enl. ed. Portland, Or: Dioscorides Press; 1992.
[52] Xu W, Qi H, Shen T, Zhao M, Song Z, Ran N, et al. Poplar coma morphogenesis and miRNA regulatory networks by combining ovary tissue sectioning and deep sequencing. iScience, 2023; 26(4): 106496.
[53] Liu W, Liu C, Jin J, Li D, Fu Y, Yuan X. High-throughput phenotyping of morphological seed and fruit characteristics using X-ray computed tomography. Front Plant Sci, 2020; 11: 601475.
[54] Han L, Wang L Y, Hu G P. A study on the machine vision assisted vibratory feeding system. Applied Mechanics and Materials, 2012; 220: 1377–1380.
[55] Han L, Deng J. A study on flexible vibratory feeding system based on halcon machine vision software. Atlantis Press, 2015; pp.9–12. Available: https://www.atlantis-press.com/article/18191
[56] Wu D, Ding D, Cui B, Jiang S, Zhao E, Liu Y, et al. Design and experiment of vibration plate type camellia fruit picking machine. Int J Agric & Biol Eng, 2022; 15(4): 130–138.
[57] Ropelewska E, Sabanci K, Aslan M F, Azizi A. A novel approach to the authentication of apricot seed cultivars using innovative models based on image texture parameters. Horticulturae, 2022; 8(5): 431.
[58] Ropelewska E. The use of seed texture features for discriminating different cultivars of stored apples. Journal of Stored Products Research, 2020; 88: 101668.
[59] Medina W, Skurtys O, Aguilera J M. Study on image analysis application for identification Quinoa seeds (*Chenopodium quinoa* Willd) geographical provenance. LWT-Food Science and Technology, 2010; 43(2): 238–246.
[60] Sharma K K, Seal A. Clustering analysis using an adaptive fused distance. Engineering Applications of Artificial Intelligence, 2020; 96: 103928.
[61] Van De Velden M, Iodice D'Enza A, Markos A. Distance‐based clustering of mixed data. WIREs Computational Stats, 2019; 11(3): e1456.
[62] Deok Han G, Mansoor S, Kim J, Park J, Heo S, Yu J K, et al. A study of the morphological and geographical diversity of Korean indigenous buckwheat landraces for breeding. Journal of King Saud University-Science, 2024; 36(9): 103387.
[63] Juma I, Nyomora A, Hovmalm H P, Fatih M, Geleta M, Carlsson AS, et al. Characterization of Tanzanian avocado using morphological traits. Diversity, 2020; 12(2): 64.
[64] Han D, Hu H, Yang J, Liang X, Ai J, Abula A, et al. The ideal harvest time for seed production in maize (*Zea mays* L.) varieties of different maturity groups. J Sci Food Agric, 2022; 102(13): 5867–5874.
[65] Morales C G D P, García-De Los Santos G, Aguilar-Rincón V H, Hernández-Livera A, Escamilla-Prado E. Coffee (*Coffea arabica* L.) harvesting time and its influence on the seed quality of the Costa Rica 95 and Garnica varieties. Agro Productividad, 2022; DOI: 10.32854/agrop.v15i8.1935.