An AI-driven image processing technique to simplify the pollen measurement in common ragweed (*Ambrosia artemisiifolia* L.)

Mesterséges intelligencia által vezérelt képfeldolgozó rendszer alkalmazása ürömlevelű parlagfű (Ambrosia artemisiifolia L.*) pollenmérésére*

Jakab Máté Scherman, Marietta Petróczy, Erzsébet Szathmáry and Gábor Markó*

Department of Plant Pathology, Institute of Plant Protection, Hungarian University of Agriculture and Life Sciences, Ménesi út 44, Budapest 1118, Hungary

**Correspondence: marko.gabor3@gmail.com*

Abstract: *Ambrosia artemisiifolia* (common ragweed) is an invasive weed species that significantly impacts agriculture and public health. This study aimed to develop an automated AI-based object detection model using our annotated image recognition dataset for accurate pollen size measurement, focusing on repeatability and variability in pollen size among individuals with distinct morphological characteristics. The model can effectively streamline the traditionally labour-intensive process, achieving rapid, accurate data collection. Roboflow-based image analysis takes only milliseconds, which is significantly faster than traditional approaches, and a high repeatability index demonstrates a valid methodology for pollen analysis. The study suggests a relationship between pollen size variability and plant morphology, suggesting possible trade-offs between growth and reproduction or showing habitat-specific adaptations. Results may create valuable opportunities for plant biology or ecology, for instance, further investigation of plant-pathogen interactions and public health research. This innovative method represents a step forward in efficient pollen analysis and its integration into multidisciplinary studies.

Keywords*: automated measurement, image dataset, digital image processing, alien species, palynology, model training, artificial intelligence, machine learning*

Összefoglalás: Az ürömlevelű parlagfű (*Ambrosia artemisiifolia*) invazív gyomnövény, amelynek jól ismert a mezőgazdaságra és az egészségre gyakorolt negatív hatása. Munkánk során egy olyan mesterséges intelligencia alapú objektum felismerő modell kidolgozását tűztük ki célul, amely az általunk annotált képfelismerési adatbázist használja a pollenek méretének pontos meghatározására. Vizsgáltuk továbbá a pollenméret variabilitását eltérő morfológiai adottságokkal rendelkező egyedek között. A modell alkalmazása jelentősen leegyszerűsítette és felgyorsította a pollenmérés lassú és munkaigényes folyamatát. A hagyományos képelemző módszerekkel összehasonlítva, a Roboflow-alapú feldolgozás néhány milliszekundumra csökkenti a mérési időt. A repeatabilitás tesztelésekor, ugyanazt a pollent többször lemérve nagyon hasonló eredményt kaptunk, ami alapján kijelenthető, hogy a módszer alkalmas a pollenek méretének megbízható számszerűsítésre. Eredményeink egyedspecifikus összefüggést mutattak ki a pollenek méretében, ami mögött vélhetően trade-off kapcsolat állhat a növény növekedése és reprodukciója között. A jövőben érdemes lehet összefüggéseket keresni a morfológiai megjelenés és a pollen mennyisége és minősége között. A módszer és a belőlük származó kezdeti eredmények értékes kutatási potenciállal bír a növénybiológia vagy az ökológia területén, a növény-kórokozó kölcsönhatások további vizsgálatához, valamint a közegészségügyi vonatkozású kutatásokhoz. Ez az innovatív módszer fontos eszközként

szolgálhat a pollenek monitorozásához és más morfológiai kutatásokhoz, amely eredmények nagy érdeklődésre tarthatnak számot multidiszciplináris alkalmazhatóságuk révén.

***Kulcsszavak**: automatizált mérés, képi adatbázis, digitális képfeldolgozás, inváziós fajok, palinológia, modellképzés, mesterséges intelligencia, gépi tanulás*

1 Introduction

Ambrosia artemisiifolia (L.) is an annual invasive weed that has spread worldwide, including in many European countries (Leiblein-Wild et al., 2014). According to a survey by the Weed Science Society of America (2017), this species ranked among the top 10 of the most harmful weeds in North American crop production (Van Wychen, 2017).

Controlling common ragweed is critical for several reasons. In crop production, it negatively affects the yields of sunflower, corn, sugar beet, soybean, and cereal crops (Buttenschøn et al., 2010). Additionally, ragweed pollen is a significant public health hazard, being the main cause of allergic rhinitis and asthma (Buttenschøn et al., 2010). Studies have shown that even low pollen concentrations (5–10 grains per m^3) can provoke health issues in sensitive persons. Furthermore, in ragweed-infested areas, up to 12% of the human population suffers from respiratory diseases (Taramarcaz et al., 2005).

Pollen plays a crucial role in plant reproductive biology, so any methodical developments in quantifying their numbers, size, and other parameters represent an important methodological breakthrough. However, visual analysis of pollen images has traditionally been labour-intensive and time-consuming, requiring numerous manual steps for sample preparation and producing detailed microscopic images (Langford et al., 1990). In a study, an ImageJ software-based method proved to be a cheap, efficient, and reliable tool for pollen counting, adaptable to various pollen types and sizes (Costa & Yang, 2009). Efforts in development related to automated image processing and image-based recognition, an increasing number of image datasets are now available for analysing pollen grains (Rodrigues et al., 2015).

This study aimed to develop a rapid, automated, and user-friendly AI-based image processing method for accurately measuring pollen size. Therefore, our specific research aims were two-fold: 1) we tested the repeatability of the measuring procedures (i.e., how similar the size of two identical pollen grains after repeated measuring), and 2) we tested the variability of the pollen size among individuals. We hypothesised that if there is an individual-specific morphological trait, we will find differences in pollen sizes among individuals with different morphological appearances. Therefore, we selected common ragweed individuals exhibiting extremely different morphological characteristics to explore the influence of specific plant traits on pollen size.

2 Materials and Methods

2.1 Pollen collection

First, common ragweed individuals (N = 3) with various morphological appearances (1. Table) were selected for pollen collection. The samples were obtained from the same maize field population near Pilisvörösvár, Hungary (GPS: 47.627, 18.926; September 23, 2024). The collected plant material (i.e., inflorescence of male flowers) was transported in paper bags. Inflorescences were carefully separated and air-dried at room temperature (20-22 °C) to avoid tissue degradation. Other plant parts were dried in a drying oven (180 °C) for 2 hours. Subsequent laboratory measurements involved weighing the dry matter of flowers and plant parts.

Table 1 The main morphological characteristics of the studied common ragweed plants

Plant ID	longest shoot height (cm)	shoot radius (cm)	stem diameter (mm)	inflorescence number	plant biomass (g)	flower biomass (g)	pollen grains*
605	67	12	4.74	74	13.03	2.75	2.81
635	115	28	6.16	65	30.28	3.43	1.5
665	33	5	2.48	9	1.05	0.19	0.53

*1. Note *Number of pollen grains extracted from 0.15g dry male flower mass, expressed in millions*

Pollen was extracted following the method of Vaudo et al. (2020). Male reproductive parts of the flowers (0.15 g) were placed in a plastic sample collector and vortexed (20 s). Due to electrostatic forces, the charged pollen grains settle on the wall of the plastic centrifuge tube, allowing for easy separation from the remaining flower material. An aqueous pollen suspension was prepared by mixing ragweed pollen with sterilised distilled water (1 mL). The collected pollen samples were similar in size and visual appearance to the reference pollen grains images in the Palynological Database (Sam et al., 2020).

2.2 Image dataset, model training and pollen measuring

Pollen grains were placed into a Bürker chamber (standard environment for scaling the dimensions) and observed by a light microscope (Leitz LABORLUX S). The images were captured at 5× optical magnification with a Moticam 1080 HDMI & USB output microscope camera using MIDSDevices software.

Roboflow (Dwyer et al., 2024) was chosen as an end-user-friendly end-to-end computer vision platform for developing and running our Ambrart_pollends object recognition model. During the data annotation phase, images of pollen grains were labelled by annotating 49 images (pixel ratio: 1920×1080, size: 2.07 MB) with 2290 pollen grains using the smart polygon annotation feature. During the model training phase, the Roboflow 3.0 Object Detection (Fast) (ROD, hereafter) model was trained on the annotated dataset named Ambrart_pollends. The model was trained successfully three times, with increasing efficiency after each cycle (Table 2, Figure 1).

Table 2 The training process of Ambeart_pollends pollen recognition by Roboflow 3.0 Object Detection (Fast) model

ambrart_pollends versions	Total images	Train images	Valid images	Test images	mAP (%)	Precision (%)	Recall (%)
(1st)	7	3	2	2	53.80	59.70	46.60
(2nd)	27	21	4	2	77.40	85.30	52.80
(3rd)	34	27	5	2	98.20	93.00	97.10

2. Note mAP (Mean Average Precision); Precision (correct predictions); Recall (successfully identified relevant labels)

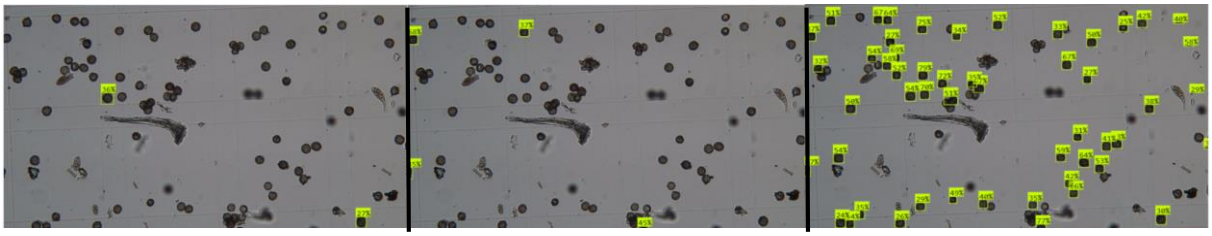

Figure 1 *The model development process demonstrates increasing effectiveness in recognition, as illustrated in the same image. (left: 1st version; middle: 2nd version, right:3rd version; see more details in Table 2.)*

After the training process, we measured the pollen regarding the focal plants by the trained model (2. Figure). The variables (position, width, height, recognition confidence, identification label) were extracted from all the recognised pollen grains and used for further statistical analyses. To determine the true size of pollen, the measurement was rescaled by the pixel µm ratio (median = 2.075003, SD = 0.0308, N = 10). For each pollen, we calculated the pollen size (ellipsoid area expressed in µm^2) using the following formula: $area = width \times height \times \pi$.

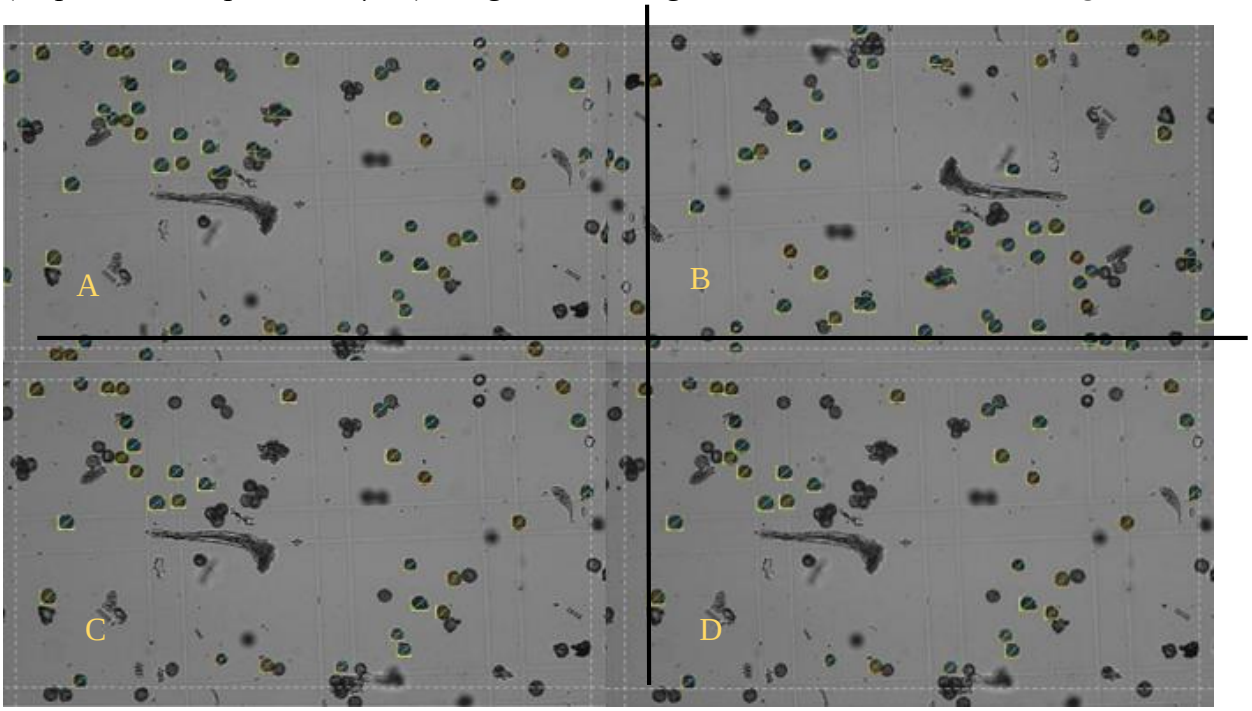

Figure 2 *Pollen recognition and measuring of the original (A, C) and rotated (B, D) versions before (A, B) and after (C, D) filtering out the false recognitions and measuring. Figure descriptions: red circle: identified pollen grains (correct recognitions and measures); blue circle: pollen grains with high asymmetry calculated from the width and height ratio (i.e., <0.9 or 1.1<; excluded from statistical analyses); yellow borders: measured width and height of given pollen grain with high recognition confidence (0.5<); orange borders: measured width and height of given pollen grain with low recognition confidence (<0.5); green line: diagonal of the fitted box; and grey dashed lines indicate the image centre (pollen grains were lying in the centre area (and without crossing these lines) were included into the statistical analyses. The recognition of the pollen grains was not completely identical comparing the original image and its rotated version.*

2.3 Statistical analyses

All statistical analyses were performed in the R statistical environment (version 4.3.1, 08-12-2024, R Development Core Team 2019).

We calculated the repeatability index by taking the ratio of within-group variance to total residual variance applied Markov Chain Monte Carlo sampler for multivariate Generalized Linear Mixed Models (MCMCglm) based on repeated measurements of the identical sample units (i.e., 'pollen_id'). The calculations were based on 1000 permutations using the '*MCMCglmm*' package (Hadfield 2010).

For testing the individual-specific variations of the pollen size, we fitted a linear mixed-effect model (LMM) structure using the '*lme4*' package (Bates et al. 2015), and the predictors were tested for significance using the Likelihood Ratio Test (LRT). In this model, the response variable was the pollen size, while the recognition confidence ('confidence', numeric, continuous), image type ('image set', factor, original or rotated, represented the repeated measuring), and plant identifier ('plant id', factor) testing the plant-specific differences. The pollen identifier ('pollen_id', factor) was included as a random variable, considering the data of the exact pollen grain from different images (i.e., original and rotated).

Table 3 Descriptive statistics of the pollen sizes and recognition confidence in each image set

Image set	Variable	Sample size	Range	Median	SD
1st	Area	101	281.23/502.89	369.24	31.21
1st	Confidence	101	0.17/0.84	0.53	0.15
2nd	Area	88	285.67/472.37	371.61	28.70
2nd	Confidence	88	0.22/0.91	0.56	0.16
Paired	Area	71 (142)	285.67/502.89	373.68	29.14
Paired	Confidence	71 (142)	0.2/0.91	0.53	0.16

3. Note The descriptive statistics suggested high similarities among the image sets. The numbers indicate unique pollen grains, while the repeated measurements of the same pollen grains are detailed in brackets.

3 Results

The three-step training process of the Ambrart_pollends pollen detection model has been applied successfully within the ROD framework. The comprehensive training for creating the training image dataset took approximately 40 work hours.

We obtained a high repeatability index (4. Table), indicating that the independent measurements of pollen grains and their rotated versions yielded highly similar, pollen-specific outputs by MCMCglm.

Table 4 Repeatability index and the corresponding Confidence Intervals (2.5 and 97.5 quartiles, CI_{95}) values by increasing the recognition confidence exclusion limits calculated by MCMCglm. These values suggested that the high overall repeatability was unaffected negatively by the data with low recognition confidence. Therefore, it was needless to exclude them from the statistical analyses

Confidence limit	R	$CI_{2.5}$	$CI_{97.5}$
0.0<	0.846	0.779	0.895
0.2<	0.847	0.778	0.897
0.3<	0.874	0.811	0.915
0.4<	0.862	0.780	0.912
0.5<	0.824	0.642	0.911
0.6<	0.809	0.106	0.958
0.7<	0.210	0.000003	0.911

The LMM revealed a significant relationship between recognition confidence and pollen size ('confidence': $F(1, 176.53) = 18.60$), $p < 0.001$), indicating that recognition confidence is positively increased with the pollen size. There is a detectable difference among plants ('plant id': $F(2, 297.40) = 3.87$, $p = 0.0219$), suggesting individual-specific, systematic variation in pollen size. In contrast, the effect of repeated measuring was unaffected ('image set': $F(1, 105.40) = 0.15$, $p = 0.7001$), suggesting that the image recognition and size measuring were unbiased by the repeated sampling sessions and produced consistent results.

4 Discussion

Our study demonstrated that the Ambrart_pollends detection model provides valid pollen size measurements, effectively streamlining a traditionally labour-intensive and time-consuming process. It can also be integrated easily with pollen counting. The developed platform reduces the need for manual measurements while allowing for the rapid collection of comprehensive pollen size data and providing accurate estimation for the concentration of a pollen suspension. Additionally, model confidence estimates can be adjusted to account for visual estimation errors. Roboflow-based image analysis is completed within milliseconds in each image, making it significantly faster than previous image-based analysis methods. Compared to Costa and Yang (2009), the analysis of individual images processed with ImageJ software ranged from 30 to 60 seconds. The high repeatability index clearly demonstrates that the AI-based measuring procedure can be used as a valid methodology for pollen measuring.

Every methodological approach has its pros and cons. One of the areas for improvement of Ambrart_pollends is the relatively long time required to train the model, as it relies on high-quality images and precise annotations of individual pollen grains. Pollen grains located at the edges of the image or overlapping with others are not always measured accurately and must be filtered out to obtain reliable pollen size data. Additionally, the costs of model training and financial maintenance could be a limiting factor.

Pollen size is related to the morphological properties of plants, though a large amount of data is needed to confirm the impact of these morphological differences. If pollen size can be used to infer trade-offs between growth and reproduction which traits vary across populations or habitats, it could provide valuable insights in plant biology and ecology. It is known that male reproductive investment in *A. artemisiifolia* increases with plant height, while female investment does not, indicating a male-biased sex allocation in taller plants (Nakahara et al., 2017). This suggests that height directly influences male reproductive success. The applications of our results extend beyond weed control and respiratory public health problems (Buttenschøn et al., 2010; Taramarcaz et al., 2005) or to the studies related to plant-pathogen interactions (Kocsis et al., 2022). Pollen grains significantly enhance the germination of plant-parasitic fungal spores, which effect is influenced primarily by the size of the pollen, as well as the specific plant taxon.

5 Conclusions

This study emphasized an efficient and accurate AI-based method for measuring ragweed pollen size, significantly reducing the human workload. With further development, this method has significant potential in plant biology, offering insights into growth-reproduction trade-offs and population- or habitat-specific variations.

Acknowledgements

This work was supported by the Research Excellence Programme of the Hungarian University of Agriculture and Life Sciences (G. Markó).

References

Bates, D., Mächler, M., Bolker, B. and Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. **67** (1) 1–48. https://doi.org/10.18637/jss.v067.i01

Buttenschøn, R., Waldispühl, C. B. 2010. Guidelines for management of common ragweed, Ambrosia artemisiifolia. http://www.EUPHRESCO.org

Costa, C. M. and Yang, S. 2009. Counting pollen grains using readily available, free image processing and analysis software. *Annals of Botany*. **104** (5) 1005–1010. https://doi.org/10.1093/aob/mcp186

Dwyer, B., Nelson, J., Hansen, T. et al. 2024. *Roboflow (Version 1.0) [Software]*. In https://roboflow.com

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*. **33** (2) 1–22. https://doi.org/10.18637/jss.v033.i02

Kocsis, I., Petróczy, M., Takács, K. Z. and Markó, G. 2022. Stimulation role of pollen grains in the initial development of Botrytis cinerea: The importance of host compatibility, cultivation status and pollen size. *Journal of Phytopathology*. **170** (11–12) 828–837. https://doi.org/10.1111/jph.13149

Langford, M., Taylor, G. E. and Flenley, J. R. 1990. Computerized identification of pollen grains by texture analysis. *Review of Palaeobotany and Palynology*. **64** (1) 197–203. https://doi.org/10.1016/0034-6667(90)90133-4

Leiblein-Wild, M. C., Kaviani, R. and Tackenberg, O. 2014. Germination and seedling frost tolerance differ between the native and invasive range in common ragweed. *Oecologia*. **174** (3) 739–750. https://doi.org/10.1007/s00442-013-2813-6

Nakahara, T., Fukano, Y., Hirota, S. K. and Yahara, T. 2018. Size advantage for male function and size-dependent sex allocation in Ambrosia artemisiifolia, a wind-pollinated plant. *Ecology and Evolution*. **8** (2) 1159–1170. https://doi.org/https://doi.org/10.1002/ece3.3722

Rodrigues, C., Barbosa Goncalves, A., Silva, G. and Pistori, H. 2015. Evaluation of Machine Learning and Bag of Visual Words Techniques for Pollen Grains Classification. *IEEE Latin America Transactions*. **13** 3498–3504. https://doi.org/10.1109/TLA.9907

Sam, S., Halbritter, H. and Heigl, H. 2020. *Ambrosia artemisiifolia* PalDat - A palynological database. https://www.paldat.org/pub/Ambrosia_artemisiifolia/304617

Taramarcaz, P., Lambelet, B., Clot, B., Keimer, C. and Hauser, C. 2005. Ragweed (Ambrosia) progression and its health risks: Will Switzerland resist this invasion? *Swiss medical weekly*. **135** 538–548. https://doi.org/10.4414/smw.2005.11201

Van Wychen, L. 2017. 2017 Survey of the Most Common and Troublesome Weeds in Broadleaf Crops, Fruits, and Vegetables. https://wssa.net/wp-content/uploads/2017-Weed-Survey_Grass-crops.xlsx

Vaudo, A. D., Tooker, J. F., Patch, H. M., Biddinger, D. J., Coccia, M., Crone, M. K., Fiely, M., Francis, J. S., Hines, H. M., Hodges, M., Jackson, S. W., Michez, D., Mu, J., Russo, L., Safari, M., Treanore, E. D., Vanderplanck, M., Yip, E., Leonard, A. S. and Grozinger, C. M. 2020. Pollen Protein: Lipid Macronutrient Ratios May Guide Broad Patterns of Bee Species Floral Preferences. *Insects*. **11** (2). 132. https://doi.org/10.3390/insects11020132