Development of a System for the Automated Identification of Herbarium Specimens with High Accuracy

Masato Shirai

University of Shimane

Atsuko Takano

Museum of Nature and Human Activities Hyogo

Takahide Kurosawa

Fukushima University

Masato Inoue

The Shimane Nature Museum of Mt. Sanbe,

Shuichiro Tagane

Kagoshima University

Tomoya Tanimoto

University of Shimane

Tohru Koganeyama

Alpha Hydraulic Engineering Consultants Co., Ltd

Hirayuki Sato

Alpha Hydraulic Engineering Consultants Co., Ltd

Tomohiko Terasawa

Alpha Hydraulic Engineering Consultants Co., Ltd

Takehito Horie

Alpha Hydraulic Engineering Consultants Co., Ltd

Isao Mandai

T.R. Workers Co., Ltd.,

Takashi Akihiro (✉ takashiakihiro@gmail.com)

University of Shimane

Research Article

Keywords: Development of a system, identification, herbarium specimens, high accuracy, dried plants, paper, mounted, plant taxonomists, accuracy

Posted Date: January 24th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1266904/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Herbarium specimens are dried plants mounted to paper. They are used by a limited number of researchers, such as plant taxonomists, as a source of information on morphology and distribution. Recently, digitised herbarium specimens have begun to be used in comprehensive research to address broader issues. However, some specimens have been misidentified, and if used, there is a risk of drawing incorrect conclusions. In this study, we successfully developed a system for identifying species names with high accuracy using an image recognition system. We developed a system with an accuracy of 96.4% using 500,554 specimen images of 2,171 plant species in 192 families that grow in Japan. It is clarified where AI is looking and make decisions, and which species is mistaken for which species. As the system can be applied to digitalised images worldwide, it should be useful for selecting and correcting misidentified herbarium specimens.

Background

Herbarium specimens were first collected about 500 years ago¹, and approximately 380 million specimens are stored in approximately 3,000 museums globally². Herbarium specimens have long been used by a limited subset of researchers, such as plant taxonomists, as the type of scientific names or the vouchers of researches, or as a source of information on morphology and distribution. However, research in the field of museomics³, in which DNA, proteins, metabolite, radio isotopes⁴, heavy metals⁵ are extracted from specimens, has recently become prevalent. As the digitisation of label data (species name, collection location, collection date, images, etc.) related to plant specimens has progressed, the data have been accumulated in international databases such as Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/>) and Integrated Digitized Biocollections (iDigBio, <https://www.idigbio.org/>). Specimen images have become big data, and can be used more freely by anybody to study, e.g., the effects of climate change by examining the flowering season of certain plant species⁶. They are also beginning to be used in comprehensive research to address imminent social issues related to conservation and food security⁷⁻⁹. However, there are problems that must be resolved for the future development of these studies. One of the major problems is the data contain misidentified specimens¹⁰⁻¹². Goodwin et al.¹² reported that at least 58% of the 4,500 specimens of African gingers had the wrong name prior to a recent taxonomic study. It is difficult for non-taxonomists to notice misidentified specimens and their presence is likely to result in analyses using incorrect data. In studies that deal with big data, the amount of labour required to check for misidentifications is enormous. Misidentified specimens need to be found and corrected quickly to ensure the value of collections as a data set, but taxonomists are unevenly distributed and too few to re-examine whole specimens for its identification. There is an urgent need to develop an artificial intelligence (AI)-based plant identification system with high accuracy.

Florian Schroff et al.¹³ developed a system that can judge a person's face. It distinguishes 8 million people from about 200 million images, and its accuracy is 99.63%¹³. Recently, various researches using

deep learning technology to determine species names from specimen images has been actively conducted^{14–15}. In 2017, Carranza-Rojaset al.¹⁴ constructed a semi-automatic identification system using 113,205 images of 1,000 species obtained from the iDigBio portal. GoogLeNet (InceptionV1) was used for the analysis, and the accuracy was 70.3%. In 2018, 253,733 images of 1,191 species obtained from the iDigBio portal were analysed using GoogLeNet, and the accuracy was 63.0%¹⁵. In 2019, the Herbarium Challenge was held using 46,469 images of 683 melastome species (Melastomataceae) provided by the New York Botanical Garden, and the winning team used SeResNext-50, SeResNext-101, and ResNet-152, with an accuracy of 89.8%².

In this study, we aimed to develop a highly accurate identification system using digitised specimen images of 2,171 plant species. We investigated the optimum number of specimens per species for improving the accuracy and found that, if we exclude specimens without leaves or large or many holes in the leaves, the accuracy would increase. In addition, we investigated which species were mistaken for other species and what part of the image was focused on when making the identification.

Results

Improvement of data sets and identification accuracy

The targets of this study were species growing in Japan. In addition to images (about 290,000) digitised by the authors using a scanner (Fig. 1a1)¹⁶ or camera (Fig. 1a2)¹⁷, approximately 260,000 specimen images were downloaded from the database (Fig.2a). Finally, 546,184 images of 3,114 species in 219 families were obtained. Images of specimen were collected from throughout Japan (Fig. 2b). The numbers of specimens varied depend upon species; 838 (27%) species had ≤ 50 specimens, 742 (24%) had ≤ 40 specimens, 539 (17%) had ≤ 30 specimens, and 108 (3%) had ≤ 20 specimens.

Using the collected images, a plant species identification system was developed. For the experiments, Inception-ResNet-v2 was used, as it is the most accurate function in pre-trained deep neural network¹⁸. The results of the first experiment showed an accuracy of 92.3% (Table 1; Supplementary Data 1). There were 319 species with average macro f-scores ≤ 0.6 , calculated as $2 \times (\text{precision} \times \text{recall}) / (\text{precision} + \text{recall})$, and the average number of images used in these experiments was 48 per species. To exclude these species from the analysis target in the second experiment, we decided to use only ≥ 50 images per species. In the case of an image containing multiple individuals or shoots in one sheet (Fig. 1b1), the individual of plants or shoots were cut out to increase the number of images (Fig. 1b2). The second experiment was conducted using 534,778 images from 2,191 species, and the accuracy of the results increased to 93.9% (Table 1a; Supplementary Data 2).

In the second set of experiments, 11,950 images were incorrectly identified. Among them, there were 767 (6.4%) specimens had only the twigs without leaves and flowers/fruits (Fig. 1c) and/or had large or many holes in the leaves (Fig. 1d), and they were clearly misidentified. Such images were discarded, and the third set of experiments was conducted using 500,554 specimens from 2,171 species. The accuracy in

this experiment was 96.2% (Table 1a; Supplementary Data 3). In the system developed in this study, the most probable species were extracted as the Top-1 and the Top-5. The correct answer rate of the Top-1 was 96.2%, while the correct answer rate within the Top-5 was 99.4% (Table 1a). In this experiment, the AI misidentified 5,195 images. We investigated whether the AI had actually misidentified them, or whether this was caused by the AI correctly identifying a sample that was previously misidentified. We re-identified 181 specimens in Herbarium of University Archives and Collections, Fukushima University (FKSE). As a results, at least 34 (19%) of the 181 specimens had been previously misidentified. Including the preliminary experiment, the system was constructed 15 times. We selected specimens which is misidentified with 6 times or more by AI and re-identified it. At least 32 (28%) of 113 specimens had been previously misidentified.

Subsequently, an identification system was developed focussing only on Pteridophytes (353 species). The number of specimen images per Pteridophytes species was higher than that of other plant species, averaging 578 per species (230 in the third experiment which excluded damaged and misidentified specimens). While the number of species decreased to about one-sixth, the number of specimens per species doubled. The accuracy of the results was 98.4% (Table 1a; Supplementary Data 5). The relationship between the number of images and the average macro f-score was investigated in the analysis performed on 2,171 species (the third experiment) and the analysis only on Pteridophytes; the larger the number of images used in the analysis, the higher the average macro f-score (Supplementary Fig. 1).

Inception-ResNet v2 analysis method used in this study was compared with Inception-ResNet v2_base, Inception v3, and VGG16 using the images included in the third experiment. The method used in this study (Inception-ResNet v2) was found to be the most accurate (Table 1b). The classification accuracy of Inception-ResNet v2, Inception v3, and VGG16 showed the same tendency as classification by ImageNet, and the accuracy of Inception-ResNet v2 was the highest. The proposed method adds two 4096-dimensional, fully connected layers after the average pooling of Inception-ResNet v2. Unlike Inception-ResNet v2_base, the class of target image can be predicted from a vector with more dimensions than the number of classes; thus, the prediction accuracy is improved compared to Inception-ResNet v2 (Supplementary Data 6–8).

The influence of labels, colour bars, scales, stamps, etc. in the sample image on identification

In the data set of Herbarium Challenge 2019, the labels on herbarium sheets were removed to prevent the AI from using the species name and other information written on the label². The image input size of the training data used in this study was 299 × 299 pixels. This size has the highest determination accuracy in ImageNet¹⁸. It is difficult for even humans to read the written characters in these images (Supplementary Fig. 2). The effect of labels on identification was investigated using the following method. A set of 5,000 correctly identified sample images were randomly selected, and the images were processed so that only the label of the sample images remained (Supplementary Fig. 2). Identification was then performed. The

probability of obtaining the correct answer by chance was 0.05% (2.3 images/5,000 images). Only three images were correctly identified, giving a correct answer rate of 0.06%.

In addition to the labels, some sample images contained colour bars, scales, and stamps (Fig. 1e1). To investigate the effect of these factors on identification, they were deleted from the images (Fig. 1e2) and a fourth set of experiment was conducted to investigate their effect on the identification accuracy. The accuracy was only slightly lower when using the images with the colour bars, scales, and stamps removed (Table 1a, Supplementary Data 4). From these results, it was clarified that the presence of a label, colour bar, scale, or stamp in the image does not significantly affect the accuracy of identification.

Does AI make misidentifications as humans?

We investigated whether the AI could correctly identify species that are frequently misidentified by collectors and collection managers (hereinafter referred to as experts). First, species that were frequently misidentified by experts were selected according to the records of identification history of the specimens in FKSE. Of the species stored in the FKSE that had 50 or more specimens, 17 species with a misidentification rate (number of misidentified or previously misidentified specimens/number of specimens) $\geq 15\%$ were classified as “frequently misidentified species” (Table 2). The average number of images used per species in the third set of experiment was 230, while the average number of images used per species of the 17 species was 328. The average macro f-score of the 2,171 species was 0.962, while the average value of the 17 species was 0.890. Experts often misidentified *Platanthera tipuloides* (L.f.) Lindl. as *Platanthera minor* (Miq.) Rchb.f. In addition, *Lespedeza homoloba* Nakai was frequently misidentified as *Lespedeza cyrtobotrya* Miq. or *Lespedeza bicolor* Turcz. We investigated whether the AI also misidentified them. It was found that, for all species, the AI made the same mistakes as the experts (Table 2).

We investigated whether the AI and experts tend to make the same misidentification. Although bluegrasses are morphologically similar to each other, the AI rarely misidentified *Poa nipponica* Koidz. as *Poa trivialis* L., *Poa annua* L., and *Poa pratensis* L. subsp. *pratensis*, but often as *Corydalis pallida* (Thunb.) Pers. var. *tenuis* Yatabe and as *Pilea hamaoi* Makino which we did not understand why the AI confused these species.

To clarify what kind of species selected for Top 2 by AI when AI is successfully identified, we selected the 1,022 images that the AI identified correctly. The species that the AI identified as the Top-2 were checked for the images. It was found that 443 (43.3%) of the 1,022 species were the species misidentified by experts (Table 2) (Supplementary Data 3).

The third case is willows (*Salix* spp.), dioecious trees that are difficult to be identified, even for experts¹⁹. The floral characters are important for species diagnosis, but those of the willows are small, less than 10 mm long, and each specimen contains either female or male flowers. The shape of leaves can be useful for the species recognition, but usually leaves start to emerge after anthesis. All the willows specimens were used as training data without separating the males and females, and as the image input size was

299 × 299 pixels, in which small floral organs are too small to be recognized. A cross-tabulation table was created for the recall and precision values of 17 species of willows (Fig. 3a), and it was found that willows species were often misidentified within the same genus. The Gradient-weighted Class Activation Mapping (Grad-CAM) results (an analysis method that displays the most important parts in different colours during AI identification²⁰) showed that, the AI displayed a tendency to use the inflorescences and infructescence, and some of the branches to which they were attached for identification, and then to use the entire leaves for identification after using the infructescence (Fig. 3b).

Is AI misidentifying species of the same genus?

Species misidentified by experts are mostly within the same genus. Therefore, we investigated whether the species misidentified by the AI belonged to the same genus. We examined the genera of the Top-5 species for 137,483 images that were correctly identified by the AI. All five of the Top-5 species were from the same genus (including the Top-1, correct identified species name) in 4.6% (6,293) of the cases, four in 9.5% of cases, three in 17.1% of cases, and two in 27.5% of cases. The Top-2–5 species were the different genus as the Top-1 in 41.6% of cases. Even experts sometimes mistakenly identify *Gynostemma pentaphyllum* (Thunb.) Makino (Cucurbitaceae) as *Causonis japonica* (Thunb.) Raf. (Vitaceae) or *Aruncus sioicus* (Walter) Fernald var. *kamtschaticus* (Maxim.) H. Hara (Rosaceae) as *Astilbe thunbergii* Miq. var. *thunbergii* (Saxifragaceae), which are in different families. We investigated whether the AI misidentified these species in the same way, and the same misidentifications were found (Fig. 4).

Identification of parts where AI is important for identification.

Easy-to-identify Pteridophytes include *Thelypteris acuminata* (Houtt.) C.V. Morton with its long terminal leaflet and *Polystichum tripterum* (Kunze) C. Presl with its long, cross-shaped basal pinnae. The average macro f-score of *T. acuminata* was 0.993, and that of *P. tripterum* was 0.998. Looking at the Grad-CAM analysis results, we found that the characteristic parts for each species were captured (Fig. 5a, 5b). For *T.*, 352 of the 648 images (54.3%) focused on the apical part of lamina. For *P.*, 1,320 of the 1,381 images (95.5%) focused on the long, cross-shaped basal pinnae.

Using Grad-CAM, we identified the parts of the image that are important for AI identification. First, we investigated the Grad-CAM analysis results and selected 287 images that were identified by focussing on a particular part of the image (Fig. 5c1). Subsequently, from each image two images were created; one in which the non-focal parts were cut out (Fig. 5c2) and the other in which the focal parts were removed (Fig. 5c3). These images were then analysed. The accuracy rate for the images containing only the non-focal parts decreased to 72.4%, while the correct answer rate for the images that included only the focal part was 54.0%. Although the area of the focal part was small and the area of the non-focal parts was large, the accuracy rate of images containing only the focal parts was low. After processing an image that contained two individuals in one specimen (Fig. 5d1) to produce an image containing only one individual (Fig. 5d2 & 3), the accuracy was 82%. We prepared an images in which the sample image was

halved vertically and horizontally and further divided it into four both vertically and horizontally (Fig. 5e). When tested, the accuracy decreased to 54%.

Publication of the system

The identification system we developed in this study is open on the web site for 2,171 species (http://tayousei.life.shimane-u.ac.jp/ai/index_all.php) and for only Pteridophytes (http://tayousei.life.shimane-u.ac.jp/ai/index_Pteridophytes.php). When an image file is dragged and dropped in the web site the Top-1 to Top-5 species are displayed with their probability of accuracy (Fig. 6).

Discussion

How to improve identification accuracy

In this study, it was clarified whether the method of selecting training data contributed to the improvement of the accuracy of the identification system. It was the deletion of a poor-quality specimens and the deletion of specimens with obvious misidentification from the training data that contributed to the improvement of the accuracy of the identification. This is the first study that has been done to improve the accuracy by changing the quality and quantity of the specimen. The data sets for the Herbarium Challenge 2021 (<https://www.kaggle.com/c/herbarium-2021-fgvc8/data>) contained multiple images without plants, only envelopes with seeds (3482.jpg, 3727.jpg, 6990.jpg, 2191201.jpg, 857396.jpg, etc), and no plants (1779.jpg, 572646.jpg, 485802.jpg, 103392.jpg, 1296004.jpg, etc). By deleting such data, it was thought that the accuracy of the data would be further improved. In the experiments conducted using only images of Pteridophytes, 204,174 images were used in this study. Of these, 88.2% were stored in National Museum of Nature and Science, Tokyo (TNS). Many Pteridophytes taxonomists have involved in identification of the Pteridophytes specimens at TNS during the two projects of exhaustive flora of the Japanese Pteridophytes^{21,22}, and the stored specimens have well-databased with high identification accuracy. This was considered one of the causes of the high accuracy of the Pteridophytes identification experiment. These results indicated that the accuracy can be improved by such work, which can be easily judged by the human eye. Re-identification of the misidentified specimens by AI revealed that at least 19% were misidentified. From this, it was found that in order to further improve the identification accuracy, it is necessary to improve the quality of the teacher data rather than improving the method.

Plants vary in size and shape depending on the stage and environment they grow, and some have flowers and fruits. Furthermore, different colours and morphologies are shown, depending on the method used to make the specimens²³. To further improve the accuracy, it is necessary to increase the number of specimens for species with low average macro f-scores. Although there were species with a high average macro f-score, even if the number of specimens is approximately 60 (for example, *Asplenium setoi* N. Murak. Et Seriz., *Stewartia monadelphica* Siebold et Zucc., *Scrophularia musashiensis* Bonati, and *Styrax*

shiraianus Makino), there were some species with average macro f-scores of ≤ 0.85 , even when the number of specimens was ≥ 250 (for example, *Agrostis gigantea* Roth, *Persicaria odorata* (Lour.) Soják subsp. *conspicua* (Nakai) Yonek., *Persicaria japonica* (Meisn.) Nakai ex Ohki var. *japonica*, *Abelia spathulata* Siebold et Zucc. var. *spathulata*, *Carex kiotensis* Franch. et Sav., *Cirsium tonense* Nakai var. *tonense*, *Vandenboschia kalamocarpa* (Hayata) Ebihara, *Persicaria maculosa* Gray subsp. *hirticaulis* (Danser) S. Ekman et Knutsson var. *pubescens* (Makino) Yonek., and *Arisaema japonicum* Blume). In the latter case, the AI often misidentified these species as the other species in the same genus. In other words, they were misidentified as species with similar morphologies. To solve this problem, it is necessary to use more teacher specimens for species with low average macro f-scores.

The effects of specimen labels, stamps, rulers, color bars, etc. contained in the specimens on the identification were investigated, and it became clear that the accuracy of the identification did not increase even if these were removed.

Similarities and differences between AI and human identification methods.

From previous studies, it was not clear what the AI was using to make its decisions. In this study, the Grad-CAM analysis revealed the important areas in specimen image for AI identification. As the accuracy decreased if parts of the image were removed and only a section of the plant was used for identification, or when the image was divided and the identification made using a reduced area (Fig. 5e), the AI appeared to first observe the whole plant and then add specific characteristic parts. The identification method of the AI may be similar to that performed by experts.

It was also not apparent in previous studies which species is mistaken for which species. In this study, we created a cross-table (Fig. 3a, Supplement Data 1-9) and investigated this information. As a result, it became clear that the AI made mistakes in the species of the same genus. Furthermore, it became clear that species with different genera and families but similar in their morphology were also mistaken as in the case of experts.

In the willow genus (*Salix* spp.), it was found that the identification method is different between AI and experts, because the floral parts that experts are paying attention to identify is too small. If the parts required for identification is small, it was thought that identification would be possible by preparing an enlarged image of the part and training it. In the case of willow, identification was possible at a certain level without using such a small part, so it was considered that the accuracy of identification could be improved by increasing the number of specimens to be trained.

Utilization of the system created in this study

In Japan, the number of plant taxonomists who are able to classify the plant taxon accurately is declining, and this trend is expected to continue. While the number of people who can correctly identify species is decreasing, the need for environmental investigation is increasing owing to active human activities and environmental change. Thus, it is necessary to develop technology that can help non-

experts to correctly identify species. The identification system developed in this study is a good candidate.

By constructing a system multiple times by changing the combination of training data and test data, it is possible to select particular specimens in which AI makes a mistake in identification multiple times. Since about 28% of the specimens selected in this way were misidentified and then correct the specimen data registered in GBIF, it can be said that our system is a good one for selecting the misidentified specimens and correcting the data. Herbaria and databases are full of misidentified specimens^{10–12}. The method developed in this study was considered to be effective for the correction of such specimens and the reduction of the number of erroneous data due to misidentified specimens in the database.

Methods

Digitisation of specimens and collection of digitised specimen images.

Specimens in the FKSE, Tottori Prefectural Museum, Rikuzentakata City Museum, Kagoshima University Museum, the Shimane Nature Museum of Mt. Sanbe, and Shimane University Faculty of Life and Environment Sciences were digitised using a scanner (EPSON DS-50000G, ES-7000HS or ES-10000G). The method has been described previously¹⁶. Specimens from the Museum of Nature and Human Activities, Hyogo were digitised using a camera (SONY α6500 Samyang AF 35 mm F2.8 FE, ISO 100), as described previously¹⁷. Digitalised images of TNS, College of Life Science, National Taiwan University, and Flora of Tokyo specimens were downloaded from the website (the URL is shown in Fig. 1a.). The TNS specimens were digitised using a camera, while the College of Life Science, National Taiwan University, and Flora of Tokyo specimens were digitised using a scanner (Fig.2a).

Deep learning model

A convolutional neural network is a neural network model mainly consisting of convolutional, pooling, and fully connected layers (Supplementary Fig. 3). The convolutional layer has a weight parameter called a filter. The input image is converted into a feature map by applying the filter. The pooling layer extracts representative values from a specific region and reduces the spatial size. In the fully connected layer, all the nodes in the layer are connected to each other, and each edge has an independent weight. Different convolutional neural network models can be designed depending on the composition of the layers. In recent years, models such as VGG²⁵, Inception²⁶, and ResNet²⁷ have been confirmed to be highly accurate. Inception is composed of inception blocks that integrate the results of multiple convolutional and pooling processes within a single layer. ResNet has a shortcut connection that prevents gradient loss. Inception-ResNet-v2 consists of an inception block with an added shortcut connection, and has been shown to possess high classification accuracy¹⁸. The performance of this model was evaluated using the ImageNet dataset with 1,000 different classes, and the Top-5 accuracy was approximately 95%. In this study, we used Inception-ResNet-v2 with two additional fully connected layers, with 4,096 nodes each, after average pooling, to perform classification on a dataset with a large number of classes. The output

of the first fully connected layer was normalised using Batch Normalisation. In Inception-ResNet-v2, the number of nodes after average pooling is 1,792, so if the number of classes exceeds 1,792, the probability of belonging to each class is predicted using fewer nodes than the number of classes. By adding a fully connected layer, it becomes possible to predict the probability of belonging to each class using a larger number of nodes.

Method of removing stamps, colour bars, and scales from images

In-house software was developed to remove stamps, colour bars, and scales with a priori knowledge of their shapes and colours.

Declarations

Acknowledgements

The authors thank Ms. Chikako Yokogi and Dr. S.Mahoro for their help with experiments. The authors thank Rikuzentakata City Museum, National Museum of Nature and Science, Tokyo, Flora of Tokyo and College of Life Science, National Taiwan University for the downloaded data. The authors thank the faculty of Life and Environmental Science at Shimane University for their financial assistance in publishing this report. We would like to thank Editage (www.editage.com) for English language editing.

Author contributions

T.A., T.H. and M.S. conceived and supervised the project. T.A. and M.S designed the experiments. T.A., A.T., T.K., S.T., and T.T. digitised the specimens. T.K., H.S., and T.T. processed the specimen images. M.S performed developed identification system. I.M made the Web site. A.T., T.K., and S.T. performed the evaluation of misidentification and data analysis. T.A., A.T., M.S., and T.K. wrote the paper. All the authors discussed the results and contributed to the study.

Funding

This work was partially funded by the SEI Group of the CSR Foundation (T.A.) and JSPS Kakenhi (grant number 21K06307 to S.T. and 19K06832 to A.T.).

Competing interest statement

The authors declare no competing interests.

Data availability statement

Some of the data used in this study can be downloaded from the database (Fig.1a). The processed and other data for which we have the copyrights are available upon reasonable request emailed to the corresponding author. However, data for which we do not have the copyrights are unavailable.

Competing interest statement

The authors declare no competing interests.

References

1. Stefanaki, A. et al. Breaking the silence of the 500-year-old smiling garden of everlasting flowers the En Tibi book herbarium. *PLoS ONE***14**, e0217779 (2019).
2. Tan, K.C., Liu, Y., Ambrose, B., Tulig, M. & Belongie, S. The herbarium challenge 2019 dataset. Preprint at <https://arxiv.org/abs/1906.05372> (2019).
3. Raxworthy, C. J. & Smith, B. T. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.* **11**, 1049–1060 (2021).
4. McLauchlan, K. K. et al. Thirteen decades of foliar isotopes indicate declining nitrogen availability in central North American grasslands. *New Phytol.***187**, 1135–1145 (2010).
5. Rudin, S. M., Murray, D. W. & Whitfeld, T. J. S. Retrospective analysis of heavy metal contamination in Rhode Island based on old and new herbarium specimens. *Appl Plant Sci.* **5**(1):1600108 (2017).
6. Primack, D. et al. Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. *Am. J. Bot.***91**, 1260–1264 (2004). *Am. J. Bot.***91**, 1260–1264 (2004).
7. Soltis, D. E. & Soltis, P. S. Mobilizing and integrating big data in studies of spatial and phylogenetic patterns of biodiversity. *Plant Divers.***38**, 264–270 (2016).
8. Soltis, P. S. Digitization of herbaria enables novel research. *Am. J. Bot.***104**, 1281–1284 (2017).
9. Fukaya, K. et al. Integrating multiple sources of ecological data to unveil macroscale species abundance. *Nat. Commun.***11**, 1–14 (2020).
10. Fujii, S. An examination of confidence in open data of specimens: *Cuscuta australis* (Convolvulaceae) *Jpn. J. Ecol.* **69**, 127–131 (2019).
11. Sikes, D. S., Copas, K., Hirsch, T., Longino, J. T. & Schigel, D. On natural history collections, digitized and not: a response to Ferro and Flick. *ZooKeys***618**, 145–158 (2016).
12. Goodwin, Z. A., Harris, D. J., Filer, D., Wood, J. R. & Scotland, R. W. Widespread mistaken identity in tropical plant collections. *Curr. Biol.* **25**, R1066–R1067 (2015).
13. Schroff, F., Kalenichenko, D. & Philbin, J. FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on CVPR*, 815–823 (2015).
14. Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E. & Joly, A. Going deeper in the automated identification of herbarium specimens. *BMC Evol. Biol.* **17**, 181 (2017).
15. Carranza-Rojas, J., Joly, A., Goëau, H., Mata-Montero, E. & Bonnet, P. Automated identification of herbarium specimens at different taxonomic levels, In *Multimedia Tools and Applications for Environmental & Biodiversity Informatics* (eds Joly, A., Vrochidis, S., Karatzas, K., Karppinen, A., & Bonnet, P.) 151–167 (Springer Cham., 2018).

16. Moriguchi, J. et al. Establishment of high-speed digitization method of herbarium specimen and construction of maintenance-free digital herbarium. *Bunrui* **12**, 41–52 (2011).
17. Takano, A. et al. Simple but long-lasting: a specimen imaging method applicable for small- and medium-sized herbaria. *PhytoKeys* **118**, 1–14 (2019).
18. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. *Proc. 31st AAAI Conf. Artif. Intell.* 4278–4284 (2017).
19. Ohashi, H. Salicaceae. in *Wild Flowers of Japan* **3**, pp. 186–187. (eds Ohashi, H., Kadota, Y., Murata, J., Yonekura, K. & Kihara H., Heibonsha, 2016)
20. Selvaraju, R. R. et al. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proc. IEEE Int. Conf. Comput. Vis.* 618–626 (2017).
21. Kurata, S. & Nakaike, T. eds. Illustrations of Pteridophytes of Japan, *vol. 18* (University of Tokyo Press, 1979–1997).
22. Ebihara, A. The Standard of Ferns and Lycophytes in Japan I & II (Gakken Plus, 2016–2017).
23. de Lutio, R., Little, D., Ambrose, B. & Belongie, S. The herbarium 2021 half-earth challenge dataset. Preprint at <https://arxiv.org/abs/2105.13808> (2021).
24. Kato, M. & Ebihara, A. *Endemic Plants of Japan* (Tokai University Press 2011).
25. Simonyan, K. & Zisserman A. Very deep convolutional networks for large-scale image recognition. Preprint at <https://arxiv.org/abs/1409.1556> (2015).
26. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens J. & Wojna Z. Rethinking the inception architecture for computer vision. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 2818–2826 (2016).
27. He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 770–778 (2016).

Tables

Table 1

(a)

Experiment No	Experiment name	accuracy		macro			weighted			Number of herbarium images	Number of plant species	Number of family
		Top1	Top5	recall	precision	fscore	recall	precision	fscore			
1	all images	92%	98%	0.844	0.824	0.825	0.927	0.923	0.923	546,184	3,114	219
2	w/o below 50 images species	94%	99%	0.898	0.892	0.892	0.942	0.939	0.939	534,778	2,191	192
3	w/o broken or misidentified specimen	96%	99%	0.929	0.921	0.923	0.964	0.962	0.962	500,554	2,171	192
4	w/o color-bar, stamp, scale	96%	99%	0.921	0.912	0.913	0.958	0.957	0.956	500,554	2,171	192
5	Only Pteridophytes	98%	100%	0.946	0.947	0.945	0.985	0.984	0.984	204,174	357	32

(b)

Analysis method	accuracy		macro			weighted			Number of herbarium images	Number of plant species	Number of family
	Top1	Top5	recall	precision	fscore	recall	precision	fscore			
Inception-ResNet-v2	96%	99%	0.929	0.921	0.923	0.964	0.962	0.962	500,554	2,171	192
Inception-ResNet v2_base	95%	99%	0.913	0.905	0.906	0.955	0.953	0.953			
Inception v3	95%	99%	0.909	0.901	0.902	0.953	0.951	0.951			
VGG16	91%	98%	0.843	0.828	0.829	0.918	0.913	0.913			

Table 2

	Species that are easily misidentified	No samples in test	Species misidentified in FKSE	No. in Top-1	No. in Top-2	Percentage in Top-1	Percentage in Top-2
1	<i>Platanthera tipuloides</i> (L.f.) Lindl.	51	<i>Platanthera minor</i> (Miq.) Rehb.f.	1	15	6%	37%
			<i>Pogonia japonica</i> Rehb.f.	0	1		
			<i>Platanthera sachalinensis</i> F.Schmidt	2	3		
2	<i>Poa nipponica</i> Koidz.	27	<i>Poa sphondyliodes</i> Trin.	0	0	15%	15%
			<i>Poa pratensis</i> L. subsp. <i>pratensis</i>	1	1		
			<i>Poa acrotena</i> Steud.	2	3		
			<i>Agrostis vulvata</i> Steud.	0	0		
			<i>Poa annua</i> L.	1	0		
3	<i>Poa trivialis</i> L.	47	<i>Poa sphondyliodes</i> Trin.	1	6	4%	32%
			<i>Poa pratensis</i> L. subsp. <i>pratensis</i>	0	0		
			<i>Poa acrotena</i> Steud.	0	2		
			<i>Agrostis clavata</i> Trin. var. <i>makino</i> Ohwi	0	1		
			<i>Poa hisauchii</i> Honda	0	0		
			<i>Agrostis clavata</i> Trin. var. <i>clavata</i>	0	0		
			<i>Poa nipponica</i> Koidz.	1	6		
4	<i>Poa hisauchii</i> Honda	19	<i>Poa acrotena</i> Steud.	2	1	16%	47%
			<i>Poa pratensis</i> L. subsp. <i>pratensis</i>	0	0		
			<i>Poa sphondyliodes</i> Trin.	1	0		
			<i>Poa nipponica</i> Koidz.	0	8		
5	<i>Lespedeza homoloba</i> Nakai	178	<i>Lespedeza cyrtobotrya</i> Miq.	2	6	3%	57%
			<i>Lespedeza bicolor</i> Turcz.	3	52		
			<i>Lespedeza buergeri</i> Miq.	0	43		
6	<i>Carex leucochlora</i> Bunge var. <i>cundolleana</i> (H.L.v. et Vaniot) Katsuy.	47	<i>Carex oxyandra</i> (Franch. et Sav.) Kud	0	0	17%	6%
			<i>Carex leucochlora</i> Bunge var. <i>gracillima</i> (Akiyama) Katsuy.	5	1		
			<i>Carex leucochlora</i> Bunge var. <i>leucochlora</i>	0	1		
			<i>Carex discordea</i> Boott var. <i>discordea</i>	3	1		
			<i>Carex conica</i> Boott var. <i>conica</i>	0	0		
			<i>Bidens biternata</i> (Lour.) Merr. et Sherff	2	2		
7	<i>Bidens pilosa</i> L. var. <i>pilosa</i>	13	<i>Bidens frondosa</i> L.	0	2	15%	15%
8	<i>Carex otariensis</i> Franch.	55	<i>Carex kiotensis</i> Franch. et Sav.	2	14	5%	44%
			<i>Carex forficula</i> Franch. et Sav.	1	6		
			<i>Carex heterolepis</i> Bunge	0	3		
			<i>Carex dimorpholepis</i> Steud.	0	1		
			<i>Carex alopecuroides</i> D.Don var. <i>chlorostachya</i> C.B.Clark	0	0		
			<i>Carex fernaldiana</i> H.L.v. et Vaniot	0	0		
			<i>Carex foliosissima</i> F.Schmidt var. <i>foliosissima</i>	0	0		
			<i>Carex kiotensis</i> Franch. et Sav.	2	14		
9	<i>Hydrocotyle ramiflora</i> Maxim.	20	<i>Hydrocotyle maritima</i> Honda	1	6	5%	45%
			<i>Hydrocotyle javanica</i> Thunb.	0	3		
			<i>Lespedeza cyrtobotrya</i> Miq.	3	15		
10	<i>Lespedeza bicolor</i> Turcz.	113	<i>Lespedeza homoloba</i> Nakai	9	40	13%	50%
			<i>Lespedeza buergeri</i> Miq.	3	2		
			<i>Astilbe odontophylla</i> Miq.	1	12		
11	<i>Aruncus dioicus</i> (Walter) Fernald var. <i>kamtschaticus</i> (Maxim.) H.Hara	51	<i>Astilbe thunbergii</i> (Siebold et Zucc.) Miq. var. <i>thunbergii</i>	1	18	6%	61%
			<i>Astilbe microphylla</i> Knoll	0	1		
			<i>Vaccinium japonicum</i> Miq.	1	11		
12	<i>Vaccinium hirtum</i> Thunb. var. <i>pubescens</i> (Koidz.) T.Yamaz.	99	<i>Vaccinium smallii</i> A.Gray var. <i>glabrum</i> Koidz.	5	39	6%	51%
			<i>Viola kusumoma</i> Makino	10	26		
			<i>Viola acuminata</i> Ledeb.	0	3		
13	<i>Viola grypoceras</i> A.Gray var. <i>grypoceras</i>	152	<i>Viola verecunda</i> A.Gray var. <i>verecunda</i>	2	10	11%	31%
			<i>Viola obtusa</i> Makino	5	8		
			<i>Poa pratensis</i> L. subsp. <i>pratensis</i>	0	2		
14	<i>Lolium arundinaceum</i> (Schreb.) Darbysh.	43	<i>Festuca ovina</i> L.	0	1	2%	9%
			<i>Festuca rubra</i> L. var. <i>rubra</i>	1	1		
			<i>Cardamine occulta</i> Hornem.	10	43		
15	<i>Cardamine scutata</i> Thunb.	139	<i>Cardamine tanakae</i> Franch. et Sav. ex Maxim.	5	8	11%	40%
			<i>Nasturtium officinale</i> R.Br.	0	4		
			<i>Arabis nipponica</i> (Franch. et Sav.) H.Boissieu	0	0		
16	<i>Salix idensis</i> Trautv. et C.A.Mey.	32	<i>Salix dolichostyla</i> Seemen subsp. <i>dolichostyla</i>	3	3	22%	31%
			<i>Salix miyabeana</i> Seemen subsp. <i>gymnolepis</i> (H.L.v. et Vaniot) H.Obashi et Yonek.	2	2		
			<i>Salix integra</i> Thunb.	1	0		
			<i>Salix triandra</i> L. subsp. <i>nipponica</i> (Franch. et Sav.) A.K.Skvortsov	1	5		
			<i>Salix futura</i> Seemen	0	0		
17	<i>Festuca rubra</i> L. var. <i>rubra</i>	32	<i>Lolium arundinaceum</i> (Schreb.) Darbysh.	1	0	6%	3%

Figures

Figure 1

Specimen images taken with a scanner (**a1**) and with a camera (**a2**). (**b1**) Multiple individuals in one specimen image and (**b2**) image divided into two. (**c**) Specimen images showing only branches and leaves that have fallen. (**d**) Specimen image with a many holes in a leaf eaten by insects. (**e1**) Specimen image with a scale, stamp, and colour bar and (**e2**) the image after the scale, stamp, and colour bar were removed.

Figure 2

(**a**) List of herbarium-stored specimens used in this study. (**b**) Locations of specimen archives.

Figure 3

(**a**) A cross-tabulation table was created for the recall and precision values of 17 species of *Salix* L in the third experiment. (1) *S. integra* Thunb., (2) *S. futura* Seemen, (3) *S. pierotii* Miq., (4) *S. udensis* Trautv. et C.A.Mey., (5) *S. miyabeana* Seemen subsp. *gymnolepis* (H.L.v. et Vaniot) H.Ohashi et Yonek., (6) *S. vulpina* Andersson subsp. *vulpina*, (7) *S. dolichostyla* Seemen subsp. *serissifolia* (Kimura) H.Ohashi et H.Nakai, (8) *S. vulpina* Andersson subsp. *alopochroa* (Kimura) H.Ohashi et Yonek., (9) *S. eriocataphylla* Kimura, (10) *S. japonica* Thunb., (11) *S. dolichostyla* Seemen subsp. *dolichostyla*, (12) *S. eriocarpa* Franch. et Sav., (13) *S. triandra* L. subsp. *nipponica* (Franch. et Sav.) A.K.Skvortsov, (14) *S. gracilistyla* Miq., (15) *S. caprea* L., (16) *S. chaenomeloides* Kimura, (17) *S. sieboldiana* Blume var. *sieboldiana*, (18) others. (**b**) Results of Grad-CAM analysis of *Salix integra* Thunb. and *Salix futura* Seemen. Red indicates the more important parts while blue represents less important parts.

Figure 4

(a) *Astilbe thunbergii* (Siebold & Zucc.) Miq. var. *thunbergii*, (b) *Astilbe odontophylla* Miq., (c) *Machilus thunbergii* (Siebold & Zucc.), and (d) *Lithocarpus edulis* (Makino) Nakai. Species (a) and (b) and species (c) and (d) are different plants with similar morphologies. These species are often misidentified by experts and were misidentified by AI in this study.

Figure 5

Results of Grad-CAM analysis of (**a**) *Thelypteris acuminata* (Houtt.) C.V.Morton and (**b**) *Polystichum tripterum* (Kunze) C.Presl. (**c**) In the Grad-CAM analysis, only part of the specimen was considered important (red). (**c2**) Images in which the non-focal part was cut out and (**c3**) images in which only the

focal part was cut out were created and used for Grad-CAM analysis. (**d1**) Multiple individuals in one specimen image. The image was divided into two images (**d2, 3**) and used for Grad-CAM analysis. (**e**) The sample image was halved vertically and horizontally, and further divided it into four both vertically and horizontally.

Figure 6

When an image file is dragged and dropped in the web site the Top-1 to Top-5 species are displayed with their probability of accuracy (Fig. 6).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFig1.jpg](#)
- [SupplementaryFig2.jpg](#)
- [SupplementaryFig3.jpg](#)
- [SupplementaryFile.zip](#)