Hart, Adam G ORCID: 0000-0002-4795-9986, Bosley, Hayley, Hooper, Chloe, Perry, Jessica, Sellors-Moore, Joel, Moore, Oliver and Goodenough, Anne E ORCID: 0000-0002-7662-6670 (2023) Assessing the accuracy of free automated plant identification applications. People and Nature. doi:10.1002/pan3.10460 (In Press)

RESEARCH ARTICLE

PEOPLE AND NATURE   BRITISH ECOLOGICAL SOCIETY

Assessing the accuracy of free automated plant identification applications

Adam G. Hart[1] ⬤ | Hayley Bosley[1] | Chloe Hooper[1] | Jessica Perry[1] |
Joel Sellors-Moore[1] | Oliver Moore[2] | Anne E. Goodenough[1] ⬤

[1]Department of Natural and Social Science, University of Gloucestershire, Cheltenham, UK

[2]Taylor Wildlife, Scotland, UK

Correspondence
Adam G. Hart
Email: ahart@glos.ac.uk; http://www.taylorwildlife.co.uk/

Handling Editor: Robert Fish

Abstract

1. Widely available and inexpensive mobile phone applications offer users, whether professional ecologists or interested amateurs, the potential for simple and rapid automated identification of species, without the need to use field guides and identification keys. The increasing accuracy of machine learning is well established, but it is currently unclear if, and under what circumstances, free-to-use mobile phone applications are accurate for identifying plants to species level in real-world field conditions.

2. We test five popular and free identification applications for plants using 857 professionally identified images of 277 species from 204 genera. Across all applications, 85% of images were identified correctly in the top five suggestions, and 69% were correct with the first suggestion. Plant type (woody, forbs, grasses, rushes/sedges, ferns/horsetails) was a significant determinant of identification performance for each application. For some applications, image saliency was also important; exposure and focus were not significant.

3. Applications performed well, with at least one of the three best-performing applications identifying 96% of images correctly as their first suggestion. We conclude that, subject to some caveats, free phone-based plant identification applications are valid and useful tools for those wanting rapid identification and for anyone wanting to engage with the natural world.

KEYWORDS
automated identification, Google Lens, iNaturalist Seek, LeafSnap, plant identification, PlantNet, PlantSnap, species ID

1 | INTRODUCTION

Ecological data are crucial for management and conservation, in particular for surveying and monitoring that support interventions (Goodenough & Hart, 2017). Knowing what species are present is fundamental, and so accurate species identification is vital. But, even for experts, identification can sometimes be extremely challenging (Bonnet et al., 2018). For non-professionals or inexperienced people (including trainee ecologists, interested amateurs and participants in citizen science partnerships), the need to navigate field guides and identification keys, and the terminology therein, to identify specimens to species level may be an obstacle to accurate data collection

or to engagement with the natural world (Rehorek & Shotwell, 2018; Schussler & Olzak, 2008).

Engagement with the natural world was a trend that emerged during the COVID-19 pandemic (Grima et al., 2020; Soga et al., 2021; Tree, 2020; Venter et al., 2020), with the rising public interest in nature becoming a common feature of the lockdown narrative (e.g. Bevis, 2020; Stewart & Eccleston, 2020; Tree, 2020). Rising interest could be linked to a rise in demand for information, and widely available and inexpensive mobile phone applications offer simple and rapid automated species identification using familiar smartphone interfaces (Joly et al., 2014; Jones, 2020; Kaur & Kaur, 2019). Uptake of these applications did indeed increase during the pandemic, with reported spikes in downloads of mobile phone applications linked to bird identification (Associated Press, 2020) and bird call identification (Eurekalert, 2021). Google Trend data also clearly show a spike in searches in May 2020 for "bird identification app" (Google Trends, 2021a) and "plant identification app" (Google Trends, 2021b) indicating enhanced interest in these platforms. Mobile phone applications potentially offer an easier and quicker method to identify species than using traditional field guides and identification keys (Joly et al., 2014; Jones, 2020; Kaur & Kaur, 2019). Applications may be organized in a similar manner to a traditional taxonomic key, with user-selected options narrowing down the possible species matches within a search framework. Increasingly though, applications or plug-in accessories use automated image analysis or audio scanning based on machine learning to provide rapid and user-friendly identification suggestions for the focal species (Jones, 2020; Robinson & Robinson, 2021). The development of deep learning and convolutional neural networks (LeCun et al., 2015) have driven considerable advances in automated recognition, and these developments in turn have driven the automated identification capabilities of many commercially available applications (Wäldchen & Mäder, 2018). Uptake of automated identification applications among the public has recently increased, with high download rates of mobile phone applications linked to bird identification (Associated Press, 2020; Eurekalert, 2021) and Google Trend data also clearly showing an increase in searches for "bird identification app" and "plant identification app" (Google Trends, 2021a, 2021b). However, popularity among the public, although potentially important for engagement, does not necessarily mean the applications are ecologically valid: to be useful in data collection, for example, to support conservation aims, automated identifications have to be accurate.

Plants are an ideal taxon on which to test the accuracy of automated identification, both within computational contexts and real-world scenarios in mobile phone applications because the taxonomic group is extremely diverse and because it is usually easier to obtain clear photographs of plants under field conditions than is often possible for animals. Since 2013, the LifeCLEF challenge has sought to develop and advance automated identification systems using deep learning frameworks (e.g. Goëau et al., 2018). As part of this challenge, eight different research groups contributed models to LifeCLEF2017 with the aim of comparing machine learning with human experts (Bonnet et al., 2018). Within this framework, the PlantCLEF group focused on plant identification and Bonnet et al. (2018) report that the best machine learning system was able to correctly identify 73.3% of photographs in a test sample of 216 images of 75 plant species from 58 genera representing 33 families. The systems tested were able to identify as many plants as three of the nine experts in the testing group, with the authors concluding that such systems are "highly powerful tools for modern botany". Similarly, Goëau et al. (2018) report on an experiment where 10,000 plant images were used to test 19 different deep learning systems in comparison with nine human experts: 88% accuracy was achieved by the automated systems. It is even possible to use deep learning frameworks to identify pollen on microscopes using automated image analysis with reference to an image library (Dunker et al., 2021).

Other studies have directly tested commercially available plant identification applications against human experts. August et al. (2020) developed an "AI naturalist" that searched Flickr content for images of plants. These images were then identified using the application PlantNet (styled as Pl@ntNet). For images with a high PlantNet "classification score", identification accuracy to family and genus were both >85% while species level identification reached ca 70% accuracy. Where a single plant was the focus of the image, rather than being part of a more complex image, identification accuracy was greatly improved.

Evaluation of the baseline accuracy of some multiple applications was tested by Jones (2020) using images of United Kingdom plant species selected to be "as contrasting as possible". Large differences in accuracy were found between applications, but the best applications correctly identified >50% of images to genus or better. Nine applications were tested but the sample size of species and images was small ($n = 38$, with each image of a different species), and the study did not account for image parameters such as focus (either in the optical sense or the sense in which August et al., 2020 used the term to indicate whether the floral subject was distinctive within the image). Mäder et al. (2021) tested the Flora Incognita application by asking two expert botanists to check the identification returned for 1000 randomly selected images submitted by real users. The botanists were able to assess 847 of these images and concluded that 787 (93%) were accurate. The Flora Incognita application was further tested by Pärtel et al. (2021), who found, under field conditions, that the application was 85.3% accurate, with images with reproductive organs, or with only the target plant in the image, being better identified.

While extensive tests of machine learning identification algorithms exist and confirm a high and improving level of accuracy, our knowledge of the accuracy of the plant identification applications available on mobile phones (and likely to be used in practice by ecologists and others seeking accurate identifications) is limited by small sample size (e.g. Jones, 2020) or focus on single applications (e.g. August et al., 2020; Mäder et al., 2021; Pärtel et al., 2021). Here, we test the three popular free plant identification applications available on iPhone and Android mobile phones—PlantNet, LeafSnap and PlantSnap—using 857 professionally identified images of 277 species from 204 genera.

We also tested multi-taxon natural history application, iNaturalist Seek and the generic photographic identification application Google Lens. In all cases, we quantified overall performance and then investigated the influence of plant type (woody, forbs, grasses, rushes/sedges, ferns/horsetails) and image parameters (optical focus, exposure, image saliency) on identification accuracy.

2 | METHODS

In total, 16 participants, all of whom were ecological practitioners, were asked to take photographs of plants using their mobile phones with default settings. All participants gave written consent for their images to be used in this study. Target species were any native or naturalized vascular plant species growing in the wild in the United Kingdom. We were primarily interested in testing the effectiveness of identification tools in the context of supporting ecology and conservation in the field, so we excluded "exotic" species more likely to be found in planted parks and gardens. Plants were photographed in situ without removal or disturbance. Participants were asked to adopt the mentality of taking a "record shot" (i.e. a photograph that showed the target clearly but taken relatively quickly in field conditions). Only one photograph was taken of any individual plant. Images were submitted electronically and given a unique file name. To replicate realistic use of plant identification applications, no images were pre-processed for example by cropping or adjusting brightness or contrast settings.

Each image was scored independently by four individuals (co-authors HB, CH, JP, JSM) to assess three image quality parameters: exposure (underexposed = −1; correctly exposed = 0; overexposed = +1); focus (1–3 scale with highest being best); and the dominance and clarity of the focal plant in the image, which is henceforth termed "saliency" (Borji et al., 2015; 1–3 scale with highest being best). Means were calculated from individual scores to give one value per parameter per image. Each image was also assessed by one person (co-author AEG) to record whether a flower was present in the image (first binary variable), whether a leaf or other foliage was present in the image (second binary variable) and whether a fruit was present in the image (third binary variable). For consistency, the terms adopted here are the general terms used within several of the plant identification applications, including PlantNet and LeafSnap, but botanical definitions were always used when making these assessments. Thus, "flower" included any parts of the flower structure of a flowering plant (including petals, sepals, sigma and stamens) or any part of the inflorescence of a grass, sedge or rush; "leaf" included leaf (woody plants/forbs), culm (sedges/rushes/grasses), blade (grasses), frond (ferns) and stem/branches (horsetails); "fruit" was a swollen and ripened ovary and, thus, only applied to woody species and forbs.

The focal plant species in each image was identified by at least two ecologists. This was either the original photographer, themselves an ecological professional, and one of the four co-authors HB, CH, JP, JSM; or one of these four co-authors and botanical expert

co-author OM. Each image was then run through five automated identification applications. These were selected by searching on Apple (via App Store) and Android (via Google Play Store) platforms for "plant id*". Only three plant-specific applications were in the first 10 free applications listed on both platforms on 3 September 2020, and these, thus, become the focal applications for this study. They were as follows: PlantNet (styled as Pl@ntNet); PlantSnap; and LeafSnap. We also included the natural history application, iNaturalist Seek, which can be used for photographs of any taxonomic group including plants, and Google Lens, a widely used multi-purpose generic image identification application that was freely available on both Apple and Android platforms on 3 September 2020.

Baseline analysis involved quantifying descriptive statistics to summarize the accuracy of each application and, for PlantNet, the association between identification accuracy and app-reported identification confidence (PlantNet was the only platform to offer this function using a numeric scale). We also considered the accuracy of the subset of images processed by iNaturalist Seek that the application reported as passing its internal confidence threshold for a confirmed (rather than possible) identification. PlantNet and LeafSnap both allowed the structure present in the image uploaded to be categorized as leaf, fruit, flower or bark; where appropriate this was done (bark being omitted as no images fell into this category). iNaturalist Seek allowed the location of each image to be entered as metadata to assist with automatic identification and this was done either using geolocation data in the image or entered manually at county level. Although PlantNet and iNaturalist Seek allowed multiple images of the same specimen to be uploaded, this function was not used since each image was from a separate specimen. Although it is recognized that reclassification of plant species taxonomy might affect the assessment of congenericity, taxonomic information used in this study followed Stace (2019). We also ensured consistency in interpreting output from all applications even where applications themselves differed internally (e.g. *Chamaenerion angustifolium, Chamerion angustifolium* and *Epilobium angustifolium* were treated as synonymous) to avoid introducing bias.

Then, to understand the botanical and image factors influencing identification accuracy, separate generalized linear models (GLMs) were constructed for each of the five identification applications. The dependent variable was always an ordinal score based on the identification performance of the focal application: 5 =top suggestion correct (i.e. the application correctly identified the plant to species level as its first option); 4 = second suggestion correct (i.e. the application's first suggestion was not correct, but the second suggestion was correct); 3 =third suggestion correct; 2 =fourth suggestion correct; 1 =fifth suggestion correct; 0 =correct identification was not in first five suggestions. Independent variables entered as fixed factors were as follows: plant type (woody plants, forbs, grasses, rushes/sedges or ferns/horsetails), flower present (0/1), leaf present (0/1) and fruit present (0/1). Independent variables entered as covariates were the image quality parameters of exposure, focus and saliency detailed above. All models were ordinal with a multinomial distribution and cumulative logit link function. In each case,

this was determined to be optimal based on comparison of delta Akaike's Information Criterion (ΔAIC) scores (Akaike, 1973; Hu et al., 2011) using the thresholds given by Burnham and Burnham and Anderson (2002): ΔAIC ≤ 2 = same support as the optimum; Δ AIC 3–4 = strong support; ΔAIC 5–9 = weak support; ΔAIC ≥ 10 essentially no support. No competing models using other statistically valid model distribution/function combinations were within ΔAIC of 10 of the optimum (scores for competing models are given in supplementary material). A further five models, one for each application, were run based on whether the correct genus was given (1) or not (0) as the top suggestion as the dependent variable. The independent variables were as described above and all models used a binomial distribution with a log link function. To ensure that the assumption of orthogonality was met and, thus, that multicollinearity within the independent factors was low enough not to confound the analyses of species or genus accuracy, variance inflation factors (VIFs) were calculated for the combination of independent variables used in all models. VIFs varied between 1.047 and 1.538 and were, thus, all substantially below the (relatively liberal) upper threshold score of 10 given by Myers (1990) and Field (2000).

3 | RESULTS

In total, 857 photographs of plants were submitted by ecologists for identification using PlantNet, PlantSnap, LeafSnap, iNaturalist Seek and Google Lens. Collectively, these images spanned 277 species from 204 genera. Images showed woody species (n = 162), forbs (n = 609), grasses (n = 51), rushes and sedges (n = 8), and ferns and horsetails (n = 27). Overall, 601 images showed one organ (flower = 298; leaf = 302; fruit = 2) with the remaining 256 images featuring multiple organs (flower/leaf = 236; fruit/leaf = 17; flower/fruit = 2). Images were generally correctly exposed (−1 to +1 with 0 being optimal: mean = −0.008±0.087 SD), well-focused (1 to 3 positive scale: mean = 2.801±0.308) and with good saliency (1 to 3 positive scale: mean = 2.831±0.266).

On average, across all applications, the first identification suggestion was correct for 69% of images, while the correct identification was in the first five suggestions for 85% of images. However, there was a substantial variation in accuracy between the applications (Figure 1). The general application, Google Lens, had a lower accuracy than the combined average of the plant-specific applications when considering the accuracy of the top suggestion (57% vs. 73%), with the non-plant-specific natural history application iNaturalist Seek being intermediate between these as 66%. However, when it came to whether the correct identification was in the first five suggestions, PlantNet, iNaturalist Seek and LeafSnap all performed consistently (95%, 93% and 92%, respectively), whereas Google Lens and PlantSnap were considerably lower (74% and 71%, respectively). However, 13 images were correctly identified on Google Lens that had not been correctly identified by any of the plant-specific applications. Overall, 19% of images were correctly identified by all
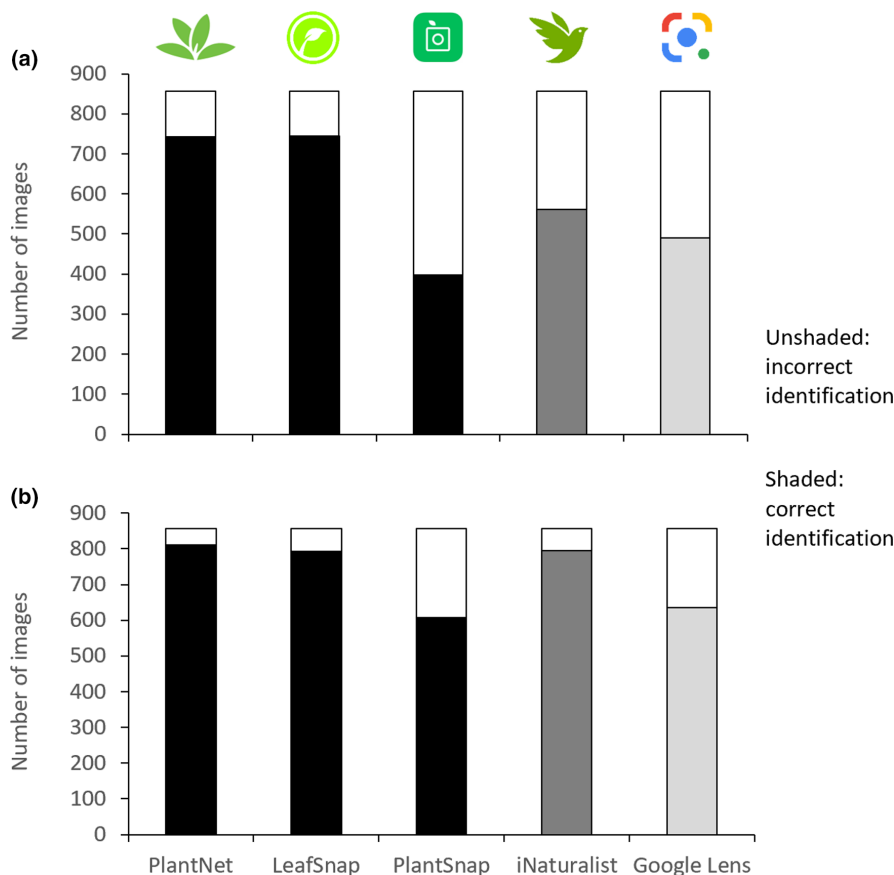

FIGURE 1 Identification accuracy for three plant-specific identification applications (black), a non-plant-specific natural history application (dark grey) and a generic image identification application (light grey) showing (a) occasions when correct identification was the first suggestion and (b) occasions when the correct identification was within top five suggestions.

five applications as the first suggestion, and when the top three performing applications, PlantNet, iNaturalist and LeafSnap, were used in triplicate, the correct identification was given as the first suggestion by at least one of the applications for 96% of images.

One application, iNaturalist Seek, provided a binary classification on whether the user should have confidence in the identification (identification confirmed) or not (multiple potential identifications suggested in order of likelihood). Of the 562 of 857 images (66%) where a confirmed identification was given, this was correct in all cases. In other words, there were no false positives within the subset of images where the application-reported confidence threshold was met. Another application, PlantNet, provided a confidence level for every identification suggestion for every image processed. At the time of the study, this was given on a 0–5 scale to two decimal places (this has now been replaced by a more-intuitive integer-based percentage but the underpinning premise remains unchanged). On average, the confidence score for a correctly identified species was 3.39 out of 5; this compared to 1.82 out of 5 for an incorrectly identified species. However, there were substantial differences between plant types (Table 1). Generally, the application was less confident of identification accuracy, even when the identification was correct, for rushes/sedges and grasses relative to forbs and woody species; confidence of correct identification for ferns and horsetails was high given the number of fern species that are morphologically similar.

The five GLMs were all significant overall (Table 2). Except for iNaturalist Seek, plant type was a significant determinant of identification performance for each application in differing ways and these patterns are explored in more detail in Figure 2. For the high-performing plant-specific applications (PlantNet and LeafSnap) none of the botanical and image variables were significant. Conversely, the performance of the high-performing but more general iNaturalist Seek was significantly higher when a flower was present in the image. The low-performing applications (PlantSnap and Google Lens) performed significantly better when a flower was present and

TABLE 1 Association between the application-reported confidence of PlantNet on a 0–5 scale for correctly identified and incorrectly identified species according to plant type.

Plant type	Confidence for the first suggestion where that suggestion is correct Mean	SD	Confidence for the first suggestion where that suggestion is incorrect Mean	SD
Woody species	3.58	1.17	0.41	0.42
Forbs	3.84	1.06	0.62	0.55
Grasses	3.01	1.33	0.92	0.52
Rushes and sedges	2.41	1.81	0.37	0.01
Ferns and horsetails	3.73	1.09	0.21	0.23

TABLE 2 Generalized linear models of factors affecting identification accuracy of each application to species level. Dependent variable gave the position of correct identification in the first five suggestions: 5 = top suggestion correct to 1 = fifth suggestion correct; 0 = correct identification did not appear in the first five suggestions. Models used a multinomial distribution and cumulative logit link function; test statistic for full model = χ^2 likelihood ratio; for separate variables χ^2 = Wald. Significant p values are shown in bold. The direction of significant results is given for simple relationships; for the effect of species type and post-hoc results, see Figure 2.

|	PlantNet χ^2	p	PlantSnap χ^2	p	LeafSnap χ^2	Dir	p	iNaturalist Seek χ^2	Dir	p	Google Lens χ^2	Dir	p
Full model	45.28	**<0.001**	45.26	**<0.001**	117.49		**<0.001**	94.67		**<0.001**	34.36		**<0.001**
Plant type	26.93	**<0.001**	24.65	**<0.001**	29.13		**<0.001**	6.67		0.154	13.03		**0.011**
Flower present?	2.30	0.129	2.40	0.121	10.94	+	**0.001**	8.59	+	**0.003**	4.43	+	**0.035**
Leaf present?	0.36	0.547	0.38	0.537	1.22		0.269	0.82		0.364	0.24		0.626
Fruit present?	0.87	0.352	0.65	0.421	2.31		0.129	0.81		0.368	1.63		0.202
Exposure	1.46	0.227	3.60	0.058	0.34		0.562	1.88		0.171	1.61		0.204
Focus	0.07	0.798	0.67	0.414	3.03		0.102	0.62		0.430	0.79		0.374
Saliency	1.21	0.271	2.37	0.124	28.27	+	**<0.001**	0.35		0.553	7.24	+	**0.007**
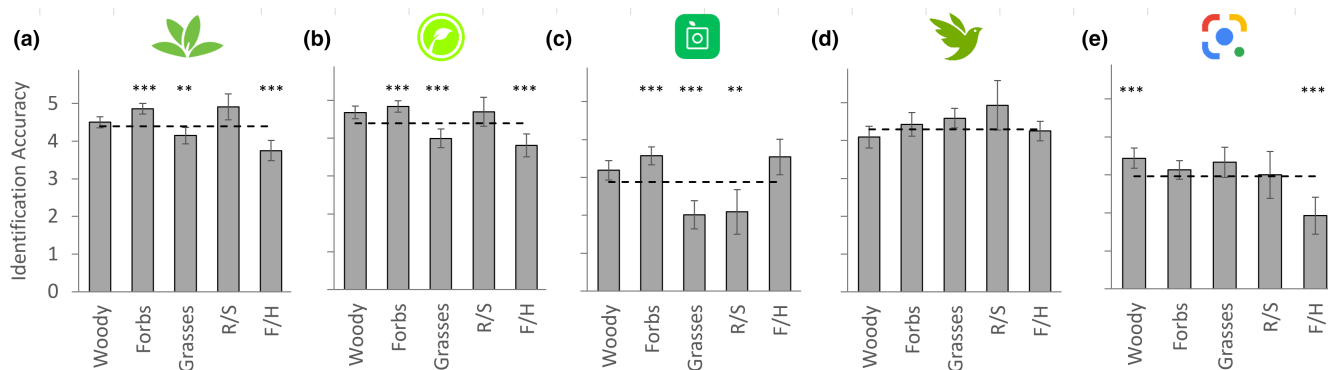
FIGURE 2 Estimated marginal means with standard error bars showing identification accuracy for plant type: (a) PlantNet; (b) LeafSnap; (c) PlantSnap; (d) iNaturalist Seek and (e) Google Lens. In all cases, the application-specific mean is shown using a dashed horizontal line and asterisks denote significant differences relative to this (** significant at α = 0.01, *** significant at α = 0.001). R/S = rushes and sedges; F/H = ferns and horsetails. Identification accuracy (y-axis): 5 = top suggestion correct (i.e. the application correctly identified the plant species as its first option); 4 = second suggestion correct (i.e. the application's first suggestion was not correct, but the second suggestion was correct); 3 = third suggestion correct; 2 = fourth suggestion correct; 1 = fifth suggestion correct; 0 = correct identification was not in first five suggestions.

where the focal plant had high saliency (i.e. the dominance and clarity of the focal plant in the image were good).

Plant type was a significant predictor variable in the GLM for all applications except iNaturalist Seek (Table 2). To explore this further, estimated marginal means (EMMs) from the GLMs were plotted and post-hoc testing was undertaken to explore significant differences in identification accuracy relative to the application-specific EMM (Figure 2). Use of EMMs, rather than raw data, allowed underlying differences in the presence/absence of organs or image quality parameters, which co-varied with plant type, to be accounted for. Identification accuracy for woody species was similar to the mean identification accuracy for each of the plant-specific applications but was significantly higher than the mean accuracy for Google Lens. Conversely, for forbs, identification accuracy on Google Lens was not significantly different from the overall accuracy found across all plant types for that application, whereas identification accuracy for forbs on the three other applications was lower than the performance of those applications across all other plant types. For the other plant types, the situation was more complex. Grass identification accuracy was significantly below mean identification accuracy for the three plant-specific applications (in other words, they performed poorly for grass identification) but not for Google Lens where accuracy did not significantly differ for grasses. Rushes and sedges were typically identified with an accuracy that did not differ the mean identification accuracy for the application, except by PlantSnap where accuracy for rushes and sedges was significantly below the PlantSnap overall average. Ferns and horsetails were typically identified with below-average accuracy, except by PlantSnap where accuracy for these plants did not differ from the mean PlantSnap accuracy. Overall, identification accuracy of the lowest-scoring plant types on PlantNet and LeafSnap was still higher than the highest-scoring plant types on PlantSnap and Google Lens. For iNaturalist Seek, there were no significant differences in identification accuracy between plant types, with accuracy being close to the

(high) mean in all cases suggesting that this application has good performance for all common plant types.

The overall accuracy at genus level (i.e. genus correctly identified as the top suggestion) was as follows: PlantNet = 97%, LeafSnap = 95%; PlantSnap = 17%; iNaturalist Seek = 93%; Google Lens = 72%. Perhaps more importantly, for those images where the correct *species* was not the first suggestion, the first suggestion was a different species in the correct genus for 65 of 115 images in PlantNet (57%), 48 of 112 images in LeafSnap (43%), 146 of 459 images in PlantSnap (32%), 238 of 297 in iNaturalist Seek (80%), and 128 of 367 images for Google Lens (35%). The factors driving the differences in accuracy at genus level between applications are shown in Table 3. Patterns largely mirror those found for identification accuracy at species level (Table 2; Figure 2), but it is notable that models the two highest performing plant-specific applications, PlantNet and LeafSnap, were non-significant, which likely reflects the very high accuracy of these applications at genus level. For iNaturalist Seek, plant type, which was non-significant when species accuracy was tested, became significant when the genus was considered; this was driven by a 100% success rate in identifying species to genus for sedges/rushes and ferns/herbs.

4 | DISCUSSION

Using 857 professionally identified images of 277 species the applications tested performed as well as, or better than in previous studies (e.g. August et al., 2020; Jones, 2020; Pärtel et al., 2021). Almost 7 out of 10 images were correctly identified by one or more applications as the first suggestion, rising to more than 9.5 out of 10 when the three best-performing applications were used together. Even when the correct identification was not the first suggestion, the correct identification, or at least the correct genus, frequently occurred within the top five suggestions. There were pronounced

TABLE 3 GLMs of factors affecting the performance of each application on whether the correct genus was identified as the top suggestion (1) or not (0). Models used a binomial distribution with a log link function; test statistic for full model = χ^2 likelihood ratio; for separate variables χ^2 = Wald. Significant p values are shown in bold.

	PlantNet			PlantSnap			LeafSnap			iNaturalist Seek			Google Lens		
	χ^2		p	χ^2	Dir	p	χ^2		p	χ^2		p	χ^2	Dir	p
Full model	17.41		0.066	98.62		**<0.001**	18.01		0.055	32.03		**<0.001**	43.38		**<0.001**
Plant type	6.30		0.178	21.15		**<0.001**	2.90		0.574	10.82		**<0.001**	21.76		**<0.001**
Flower present?	1.36		0.243	26.09	+	**<0.001**	4.38		0.066	2.70		0.101	4.41	+	**0.036**
Leaf present?	0.21		0.650	0.16		0.668	1.97		0.1260	0.01		0.916	1.21		0.252
Fruit present?	0.00		0.999	2.10		0.147	0.15		0.703	3.05		0.081	0.23		0.629
Exposure	0.824		0.364	1.01		0.314	1.86		0.173	2.33		0.127	0.00		0.984
Focus	0.705		0.401	3.13		0.069	0.26		0.609	0.00		0.983	0.16		0.693
Saliency	1.04		0.307	10.41	+	**0.001**	2.28		0.131	0.13		0.717	8.96	+	**0.003**

differences in identification performance between applications, but two plant-specific applications performed better for all plant types. LeafSnap and PlantNet achieved overall accuracies on their first-choice identifications of 86.9% and 86.6% respectively (cf. 73.3% accuracy achieved by the best machine learning algorithm reported by Bonnet et al. (2018)). Within all applications, the relative performance for different plant types was mixed, with typically better performance for woody species and forbs compared with lower performance for species such as grasses and ferns/horsetails, where identification accuracy often depends on small-scale features such as ligules and sporangia (Bonnet et al., 2018).

Our data clearly show that the identification accuracy of these free mobile phone applications is sufficiently high that they would be extremely useful in providing species-level identifications within applied ecological contexts. Having confidence in the identifications provided is vital if apps are to be of use. In this regard, iNaturalist was notable. This app has a confidence threshold that, if exceeded, results in a "confirmed" identification. Of 857 images, 562 images (66%) received a confirmed identification, which was correct in all cases with no false positives. Thus, although the app had slightly lower overall identification accuracy than other apps, when it was confident, that confidence was well-founded. In many professional scenarios, having this level of confidence in species identification may be more desirable than having less certain identifications. We are not suggesting that automated identification applications are yet at a stage where they could replace traditional identification methods, or the field skills of expert botanists. However, as noted by other authors (e.g. Bonnet et al., 2018), these applications are already capable of generating high-quality ecological data and therefore being useful to support ecological surveying and monitoring activities. The intuitive and accessible nature of applications (Joly et al., 2014; Jones, 2020), especially when compared with conventional identification keys (Rehorek & Shotwell, 2018), also give them the potential to support a wider user base, which might be particularly helpful in citizen science projects or for trainee ecologists in early-career positions or within educational contexts.

It should be noted that overall application performance might be overestimated in most evaluations (including here) because the images tested will, unless there is a specific selection otherwise, tend to show relatively common species that are likely to be well-represented in the images from which these applications learn. On the other hand, in our research, no image pre-processing was undertaken to optimize record shots before they were submitted to each application, which might have reduced the accuracy of identification for some images. It is possible that specific applications were more sensitive to this lack of image optimisation, with anecdotal evidence (Goodenough and Hart, pers. obs.) suggesting that Google Lens performs better when images are cropped to showcase key parts of plants in more detail.

Identification accuracy is ultimately related to the opportunities the applications have to learn, which in turn is a function of the number of images uploaded. This likely explains at least some of the accuracy differences between different plant types. For example, forbs

with attractive flowers attract more uploads than grasses, and in turn, those features leading to more attention from users are also likely to be easier to distinguish in images. Another factor limiting uploaded images, and therefore accuracy, is location. This study focused on the United Kingdom where there are likely to be far more users uploading images than in less developed, less populated or more remote locations. In such cases, identification accuracy is probably far lower overall, although may still be high for those species that are particularly prominent or have wider geographical ranges.

Identifying the plant species present at most field sites requires either a very extensive key (e.g. Stace, 2019), or multiple field guides covering not only trees, shrubs and wildflowers but also plant types that are typically excluded from general volumes, such as grasses, sedges and rushes. Even then, non-native species and garden escapes might not be included (which can also be problematic for automated identification, August et al., 2020). Carrying guides as PDFs on mobile devices saves the weight and inconvenience that physical copies impose, but does not necessarily save any time in identification. Professionals could make use of automated applications by using them to support identifications, akin to getting a second opinion, or by using them to suggest identifications for specimens outside of a practitioner's expertise. These preliminary identifications could then be followed up using specific taxonomic keys and other resources, with applications potentially saving considerable time. In this way, applications and traditional methods can work in a complementary fashion for professional users. For the inexperienced or non-professional user applications provide an excellent entry point for finding out more about plants, albeit with some limitations. By providing a tentative identification, applications can assist novice users in identifying plants either directly (when identifications are correct) or indirectly (by providing guidance as to where to look in a field guide to explore related plants). This is important as without the guidance provided by the application, inexperienced users may struggle to know where to start and could perhaps become disengaged. Likewise, the taxonomic arrangement of field guides and the guidance they provide on why a particular species is identified as such (including the key taxonomic and identification features) gives users the chance to gain knowledge and confidence, potentially enhancing engagement. As with professional users, the complementarity of the two identification approaches can provide benefits that neither can give on their own.

The way in which users interface with applications and make use of them in real-world scenarios, as well as the overall user experience, are likely to almost as important in determining the uptake of automated identification applications as their actual identification accuracy. Currently, there is little information on if, and how, the professional community makes use of this technology and how ecological practitioners view such digital aids. We also know little about how non-professionals make use of identification applications. We suggest that now the accuracy of applications has been well documented (e.g. this study; August et al., 2020; Jones, 2020), research focusing on user experience would be informative to find out more about their current and potential use in all contexts.

We conclude that free mobile phone automated applications already provide accurate identifications to an acceptable taxonomic level. Over time, application identification accuracy will further improve, with a greater number of reference images, enhanced algorithms and improved machine learning (Bonnet et al., 2018; Kaur & Kaur, 2019). Although for some species the need for microscopic examination might prove an insurmountable obstacle, the range of taxonomic groups able to be usefully identified using automated applications will also increase if demand exists to make these automated identification systems ever more attractive to all users.

AUTHOR CONTRIBUTIONS

Adam G. Hart and Anne E. Goodenough conceived the ideas, designed the methodology, analysed the data and led the writing of the paper and should be viewed as equal first authors. Oliver Moore identified images. Hayley Bosley, Chloe Hooper, Jessica Perry and Joel Sellors-Moore identified images and collected the data using mobile phone applications.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data sets generated during and/or analysed during the current study are available from the University of Gloucestershire repository at http://eprints.glos.ac.uk/10364/ (dataset) http://eprints.glos.ac.uk/10365/ (image library).

ORCID

Adam G. Hart 🔟 https://orcid.org/0000-0002-4795-9986
Anne E. Goodenough 🔟 https://orcid.org/0000-0002-7662-6670

REFERENCES

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. V. Petrov & B. F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Academiai Kiado.

Associated Press. (2020). Bird-watching takes flight amid coronavirus outbreak as Americans head back outdoors. May 2nd 2020. https://www.latimes.com/world-nation/story/2020-05-03/bird-watching-soars-amid-covid-19-as-americans-head-outdoors

August, T. A., Pescott, O. L., Joly, A., & Bonnet, P. (2020). AI naturalists might hold the key to unlocking biodiversity data in social media imagery. *Patterns*, 1, 100116.

Bevis, G. (2020). *Coronavirus: Lockdown leads to walkabout discoveries.* BBC News, 25th May 2020 https://www.bbc.co.uk/news/uk-england-nottinghamshire-52647258

Bonnet, P., Goëau, H., Hang, S. T., Lasseck, M., Šulc, M., Malécot, V., Jauzein, P., Melet, J. C., You, C., & Joly, A. (2018). Plant identification: Experts vs. machines in the era of deep learning. In A. V. Joly,

K. Karatzas, & P. Bonnet (Eds.), *Multimedia tools and applications for environmental & biodiversity informatics, multimedia systems and applications*. Springer.

Borji, A., Cheng, M. M., Jiang, H., & Li, J. (2015). Salient object detection: A benchmark. *IEEE Transactions on Image Processing*, 24(12), 5706–5722.

Burnham, K. P., & Anderson, D. R. (2002). Multimodel avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management*, 66(3), 912–918.

Dunker, S., Motivans, E., Rakosy, D., Boho, D., Mäder, P., Hornick, T., & Knight, T. M. (2021). Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist*, 229, 593–606.

Eurekalert. (2021). Bird call app downloaded one million times worldwide -- now available for IOS devices. 10th March 2021. https://www.eurekalert.org/pub_releases/2021-03/cuot-bca031021.php

Field, A. (2000). *Discovering statistics*. Sage.

Goëau, H., Bonnet, P., & Joly, A. (2018). Overview of ExpertLifeCLEF 2018: how far automated identification systems are from the best experts? In *CLEF-Conference and Labs of the Evaluation Forum (No. 2125)*. CEUR Workshop Proceedings (CEUR-WS.org).

Goodenough, A., & Hart, A. G. (2017). *Applied ecology: Monitoring, managing, and conserving*. Oxford University Press.

Google Trends. (2021a). https://trends.google.com/trends/explore?date=today%205-y&geo=GB&q=bird%20identification%20app.

Google Trends. (2021b). https://trends.google.com/trends/explore?date=today%205-y&geo=GB&q=plant%20identification%20app.

Grima, N., Corcoran, W., Hill-James, C., Langton, B., Sommer, H., & Fisher, B. (2020). The importance of urban natural areas and urban ecosystem services during the COVID-19 pandemic. *PLoS ONE*, 15, e0243344.

Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: Examples from an HIV-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse*, 37, 367–375.

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J. F., & Boujemaa, N. (2014). Interactive plant identification based on social image data. *Ecological Informatics*, 23, 22–34.

Jones, H. G. (2020). What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. *AoB Plants*, 12, plaa052.

Kaur, S., & Kaur, P. (2019). Plant species identification based on plant leaf using computer vision and machine learning techniques. *Journal of Management Information Systems*, 6, 49–60.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.

Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A., & Wäldchen, J. (2021). The flora incognita app–interactive plant species identification. *Methods in Ecology and Evolution*, 12, 1335–1342.

Myers, R. (1990). *Classical and modern regression with applications*.

Pärtel, J., Pärtel, M., & Wäldchen, J. (2021). Plant image identification application demonstrates high accuracy in northern Europe. *AoB Plants*, 13, plab050.

Rehorek, S. J., & Shotwell, M. A. (2018). How to use taxonomic principles in a non-scientific setting to teach hierarchical thinking. *The American Biology Teacher*, 80, 446–450.

Robinson, C. V., & Robinson, J. M. (2021). Listen but do not touch: Using a smartphone acoustic device to investigate bat activity, with implications for community-based monitoring. *Acta Chiropterologica*, 23(1), 247–258.

Schussler, E. E., & Olzak, L. A. (2008). It's not easy being green: Student recall of plant and animal images. *Journal of Biological Education*, 42, 112–118.

Soga, M., Evans, M. J., Cox, D. T., & Gaston, K. J. (2021). Impacts of the COVID-19 pandemic on human–nature interactions: Pathways, evidence and implications. *People and Nature*, 3, 518–527. https://doi.org/10.1002/pan3.10201

Stace, C. (2019). *New flora of the British Isles*. C&M Floristics.

Stewart, D., & Eccleston, J. (2020). Enjoying the outdoors: Monitoring the impact of coronavirus and social distancing. NatureScot Research Report No. 1252.

Tree, I. (2020). Lockdown awakened our interest in nature, but it mustn't be at the expense of wildlife. The Guardian, Monday 28th December 2020 https://www.theguardian.com/commentisfree/2020/dec/28/lockdown-nature-expense-wildlife

Venter, Z. S., Barton, D. N., Gundersen, V., Figari, H., & Nowell, M. (2020). Urban nature in a time of crisis: Recreational use of green space increases during the COVID-19 outbreak in Oslo, Norway. *Environmental Research Letters*, 15, 10407.

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9, 2216–2225.

How to cite this article: Hart, A. G., Bosley, H., Hooper, C., Perry, J., Sellors-Moore, J., Moore, O., & Goodenough, A. E. (2023). Assessing the accuracy of free automated plant identification applications. *People and Nature*, 00, 1–9. https://doi.org/10.1002/pan3.10460