

1 **Machine learning to classify animal species in camera trap images: applications in ecology**

2
3 Michael A. Tabak^{1,2}, Mohammed S. Norouzzadeh³, David W. Wolfson¹, Steven J. Sweeney¹,
4 Kurt C. VerCauteren⁴, Nathan P. Snow⁴, Joseph M. Halseth⁴, Paul A. Di Salvo¹, Jesse S. Lewis⁵,
5 Michael D. White⁶, Ben Teton⁶, James C. Beasley⁷, Peter E. Schlichting⁷, Raoul K. Boughton⁸,
6 Bethany Wight⁸, Eric S. Newkirk⁹, Jacob S. Ivan⁹, Eric A. Odell⁹, Ryan K. Brook¹⁰, Paul M.
7 Lukacs¹¹, Anna K. Moeller¹¹, Elizabeth G. Mandeville², Jeff Clune³, Ryan S. Miller¹

8
9 ¹ Center for Epidemiology and Animal Health; United States Department of Agriculture; 2150
10 Centre Ave., Bldg B, Fort Collins, CO 80526

11 ² Department of Zoology and Physiology; University of Wyoming; 1000 E. University Ave.,
12 Laramie, WY 52071

13 ³ Computer Science Department; University of Wyoming; 1000 E. University Ave., Laramie,
14 WY 52071

15 ⁴ National Wildlife Research Center; United States Department of Agriculture; 4101 Laporte
16 Ave., Fort Collins, CO 80521

17 ⁵ College of Integrative Sciences and Arts; Arizona State University; 66307 South Backus Mall,
18 Mesa, AZ 85212

19 ⁶ Tejon Ranch Conservancy, 1037 Bear Trap Rd, Lebec, CA, 93243

20 ⁷ University of Georgia, Savannah River Ecology Laboratory, Warnell School of Forestry and
21 Natural Resources, PO Drawer E, Aiken, SC 29802, USA

22 ⁸ Range Cattle Research and Education Center; Wildlife Ecology and Conservation; University of
23 Florida; 3401 Experiment Station, Ona, Florida 33865

24 ⁹ Colorado Parks and Wildlife; 317 W. Prospect Rd., Fort Collins, CO 80526

25 ¹⁰ Department of Animal and Poultry Science; University of Saskatchewan; 5 Campus Drive,
26 Saskatoon, SK, Canada S7N 5A8

27 ¹¹ Wildlife Biology Program, Department of Ecosystem and Conservation Sciences; W.A. Franke
28 College of Forestry and Conservation; University of Montana; 32 Campus Drive, Missoula, MT
29 59812

30

31

32 Running Title: Machine learning to classify animals

33

34 Word Count: 6,970 includes tables, figures, and references

35

36 Corresponding Authors:

37 Michael Tabak & Ryan Miller

38 Center for Epidemiology and Animal Health

39 United States Department of Agriculture

40 2150 Centre Ave., Bldg B

41 Fort Collins, CO 80526

42 +1-970-494-7272

43 Michael.a.tabak@aphis.usda.gov

44 Ryan.s.miller@aphis.usda.gov

45 **Abstract**

46 1. Motion-activated cameras (“camera traps”) are increasingly used in ecological and
47 management studies for remotely observing wildlife and have been regarded as among the most
48 powerful tools for wildlife research. However, studies involving camera traps result in millions
49 of images that need to be analyzed, typically by visually observing each image, in order to
50 extract data that can be used in ecological analyses.

51 2. We trained machine learning models using convolutional neural networks with the ResNet-18
52 architecture and 3,367,383 images to automatically classify wildlife species from camera trap
53 images obtained from five states across the United States. We tested our model on an
54 independent subset of images not seen during training from the United States and on an out-of-
55 sample (or “out-of-distribution” in the machine learning literature) dataset of ungulate images
56 from Canada. We also tested the ability of our model to distinguish empty images from those
57 with animals in another out-of-sample dataset from Tanzania, containing a faunal community
58 that was novel to the model.

59 3. The trained model classified approximately 2,000 images per minute on a laptop computer
60 with 16 gigabytes of RAM. The trained model achieved 98% accuracy at identifying species in
61 the United States, the highest accuracy of such a model to date. Out-of-sample validation from
62 Canada achieved 82% accuracy, and correctly identified 94% of images containing an animal in
63 the dataset from Tanzania. We provide an R package (Machine Learning for Wildlife Image
64 Classification; MLWIC) that allows the users to A) implement the trained model presented here
65 and B) train their own model using classified images of wildlife from their studies.

66 4. The use of machine learning to rapidly and accurately classify wildlife in camera trap images
67 can facilitate non-invasive sampling designs in ecological studies by reducing the burden of
68 manually analyzing images. We present an R package making these methods accessible to
69 ecologists. We discuss the implications of this technology for ecology and considerations that
70 should be addressed in future implementations of these methods.

71 Keywords: artificial intelligence, camera trap, convolutional neural network, deep learning, deep
72 neural networks, image classification, machine learning, R package, remote sensing, wildlife
73 game camera

74 **Introduction**

75 An understanding of species' distributions is fundamental to many questions in ecology
76 (MacArthur, 1984; Brown, 1995). Observations of wildlife can be used to model species
77 distributions and population abundance and evaluate how these metrics relate to environmental
78 conditions (Elith, Kearney, & Phillips, 2010; Tikhonov et al., 2017). However, developing
79 statistically sound data for species observations is often difficult and expensive (Underwood,
80 Chapman, & Connell, 2000) and significant effort has been devoted to correcting bias in more
81 easily collected or opportunistic observation data (Royle & Dorazio, 2008; MacKenzie et al.,
82 2017). Recently, technological advances have improved our ability to observe animals remotely.
83 Sampling methods such as acoustic recordings, images from crewless aircraft (or “drones”), and
84 motion-activated cameras that automatically photograph wildlife (i.e., “camera traps”) are
85 commonly used (Blumstein et al., 2011; O’Connell et al., 2011; Getzin et al., 2012). These tools
86 offer great promise for increasing efficiency of observing wildlife remotely over large
87 geographical areas with minimal human involvement and have made considerable contributions
88 to ecology (Rovero et al., 2013; Howe et al., 2017). However, a common limitation is these
89 methods lead to a large accumulation of data – audio and video recordings and images – which
90 must be first classified in order to be used in ecological studies predicting occupancy or
91 abundance (Swanson et al., 2015; Niedballa et al., 2016). The large burden of classification, such
92 as manually viewing and classifying images from camera traps, often constrains studies by
93 reducing the sampling intensity (e.g., number of cameras deployed), limiting the geographical
94 extent and duration of studies. Recently, machine learning has emerged as a potential solution for
95 automatically classifying recordings and images.

96 Machine learning methods have been used to classify wildlife in camera trap images with
97 varying levels of success and human involvement in the process. One application of a machine
98 learning approach has been to distinguish empty and non-target animal images from those
99 containing the target species to reduce the number of images requiring manual classification.
100 This approach has been generally successful, allowing researchers to remove up to 76% of
101 images containing non-target species (Swinnen et al., 2014). Development of methods to identify
102 several wildlife species in images has been more problematic. Yu et al. (2013) used sparse
103 coding spatial pyramid matching (Lazebnik, Schmid, & Ponce, 2006) to identify 18 species in
104 images, achieving high accuracy (82%), but their approach necessitates each training image to be
105 manually cropped, requiring a large time investment. Attempts to use machine learning to
106 classify species in images without manual cropping have achieved far lower accuracies: 38%
107 (Chen et al., 2014) and 57% (Gomez Villa, Salazar, & Vargas, 2017). However, more recently
108 Norouzzadeh et al. (2018) used convolutional neural networks with 3.2 million classified images
109 from camera traps to automatically classify 48 species of Serengeti wildlife in images with 95%
110 accuracy.

111 Despite these advances in automatically identifying wildlife in camera trap images, the
112 approaches remain study specific and the technology is generally inaccessible to most ecologists.
113 Training such models typically requires extensive computer programming skills and tools for
114 novice programmers (e.g., an R package) are limited. Making this technology available to
115 ecologists has the potential to greatly expand ecological inquiry and non-invasive sampling
116 designs, allowing for larger and longer-term ecological studies. In addition, automated
117 approaches to identifying wildlife in camera trap images have important applications in detecting
118 invasive species or rare species and improving their management.

119 We sought to develop a machine learning approach that can be applied across study sites and
120 provide software that ecologists can use for identification of wildlife in their own camera trap
121 images. Using over three million identified images of wildlife from camera traps from five
122 locations across the United States, we trained and tested deep learning models that automatically
123 classify wildlife. We provide an R package (Machine Learning for Wildlife Image Classification;
124 MLWIC) that allows researchers to classify camera trap images from North America or train
125 their own machine learning models to classify images. We also address some basic issues in the
126 potential use of machine learning for classifying wildlife in camera trap images in ecology.
127 Because our approach nearly eliminates the need for manual curation of camera trap images we
128 also discuss how this new technology can be applied to improve ecological studies in the future.

129

130 **Materials and Methods**

131 *Camera trap images*

132 Species in camera trap images from five locations across the United States (California, Colorado,
133 Florida, South Carolina, and Texas) were identified manually by researchers (see Appendix S1
134 for a description of each field location). Images were either classified by a single wildlife expert
135 or evaluated independently by two researchers; any conflicts were decided by a third observer
136 (Appendix S1). If any part of an animal (e.g., leg or ear) was identified as being present in an
137 image, this was included as an image of the species. This resulted in a total of 3,741,656
138 classified images that included 28 species or groups (see Table 1) across the study locations.
139 Images were re-sized to a resolution of 256 x 256 pixels using a custom Python script before
140 running models to increase processing speed. A subset of images (approximately 10%) was

141 withheld using conditional sampling to be used for testing of the model (described below). This
142 resulted in 3,367,383 images used to train the model and 374,273 images used for testing.

143

144 *Machine learning process*

145 Supervised machine learning algorithms use training examples to “learn” how to complete a task
146 (Mohri, Rostamizadeh, & Talwalkar, 2012; Goodfellow, Bengio, & Courville, 2016). One
147 popular class of machine learning algorithms is artificial neural network, which loosely mimics
148 the learning behavior of the mammalian brain (Gurney, 2014; Goodfellow et al., 2016). An
149 artificial neuron in a neural network has several inputs, each with an associated weight. For each
150 artificial neuron, the inputs are multiplied by the weights, summed, and then evaluated by a non-
151 linear function, which is called the activation function (e.g., Sigmoid, Tanh, or Sine). Usually
152 each neuron also has an extra connection with a constant input value of 1 and its associated
153 weight, called a “bias,” for neurons. The result of the activation function can be passed as input
154 into other artificial neurons or serve as network outputs. For example, consider an artificial
155 neuron with three inputs (I_1 , I_2 , and I_3); the output (θ) is calculated based on:

$$156 \quad \theta = \text{Tanh}(w_1 I_1 + w_2 I_2 + w_3 I_3 + w_4 I_b) \text{ (eqn 1),}$$

157 where w_1 , w_2 , w_3 and w_4 are the weights associated with each input, I_b is the bias, and $\text{Tanh}(x)$
158 is the activation function (Fig. 1). To solve complex problems multiple neurons are needed, so
159 we put them into a network. We arrange neurons in a hierarchical structure of layers; neurons in
160 each layer take input from the previous layer, process them, and pass the output to the next layer.
161 Then, an algorithm, called backpropagation (Rumelhart, Hinton, & Williams, 1986), tunes the
162 parameters of the neural network (weights and bias values) enabling it to produce the desired

163 output when we feed an input to the network. This process is called training. To adjust the
164 weights, we define a loss function as a measure of the difference between the predicted (current)
165 output of the neural network and the correct output (Y). The loss function (L) is the mean
166 squared error:

$$167 \quad L = \frac{1}{n} \sum_{i=1}^n (Y - \theta)^2 \text{ (eqn2).}$$

168 We compute the contribution of each weight to the loss value ($\frac{dL}{dW}$) using the chain rule in
169 calculus. Weights are then adjusted so the loss value is minimized. In this “weight update” step,
170 all the weights are updated to minimize L :

$$171 \quad w_i = w_{i \text{ initial}} - \eta \frac{dL}{dW} \text{ (eqn 3),}$$

172 where η is the learning rate and is chosen by the scientist. A higher η indicates larger steps are
173 taken per training sample, which may be faster, but a value that is too large will be imprecise and
174 can destabilize learning. After adjusting the weights, the same input should result in an output
175 that is closer to the desired output. For more details of backpropagation and training, see
176 Goodfellow et al., 2016.

177 In fully connected neural networks, each neuron in every layer is connected to (provides input to)
178 every neuron in the next layer. Conversely, in convolutional neural networks, which are inspired
179 by the retina of the human eye, several convolutional layers exist in which each neuron only
180 receives input from a small sliding subset of neurons (“receptive field”) in the previous layer. We
181 call the output of a group of neurons the “feature map,” which depicts the response of a neuron
182 to its input. When we use convolutional neural networks to classify animal images, the receptive
183 field of neurons in the first layer of the network is a sliding subset of the image. In subsequent

184 layers, the receptive field of neurons is a sliding subset of the feature map from previous layers.
185 We interpret the output of the final layer as the probability of the presence of species in the
186 image. A softmax function is used at the final layer to ensure that the outputs sum to one. For
187 more details on this process, see Simonyan & Zisserman, 2014.

188 Deep neural networks (or “deep learning”) are artificial networks with several (> 3) layers of
189 structure. In our example, we provided a set of animal images from camera traps of different
190 species and their labels (species identifiers) to a deep neural network, and the model learned how
191 to identify species in other images that were not used for training. Once a model is trained, we
192 can use it to classify new images. The trained model uses the output of the final layer in the
193 network to assign a confidence to each species or group it evaluates, where the total confidence
194 assigned to all groups for each image sums to one. Generally, the majority of the confidence is
195 attributed to one group, the “top guess.” For example, for 90% of the images in our test dataset,
196 the model attributed $> 95\%$ confidence to the top guess. Therefore, for the purpose of this paper,
197 we mainly discuss accuracy with regard to the top guess, but our R package presents the five
198 groups with the highest confidence, the “top five guesses,” and the confidence associated with
199 each guess.

200 Neural network architecture refers to several details about the network including the type and
201 number of neurons and the number of layers. We trained a deep convolutional neural network
202 (ResNet-18) architecture because it has few parameters, but performs well; see He et al. (2016)
203 for full details of this architecture. Networks were trained in the TensorFlow framework (Adabi
204 et al., 2016) using Mount Moran, a high performance computing cluster (Advanced Research
205 Computing Center, 2012). First, since invasive wild pigs (*Sus scrofa*) are a subject of several of
206 our field studies, we developed a “Pig/no pig” model, in which we determined if a pig was either

207 present or absent in the image. In the “Species Level” model, we identified images to the species
208 level when possible. Specifically, if our classified image dataset included < 2,000 images for a
209 species, it was either grouped with taxonomically similar species (by genera, families, or order),
210 or it was not included in the trained model (Table 1). In the “Group Level” model, species were
211 grouped with taxonomically similar species into classifications that had ecological relevance
212 (Appendix S2). The Group Level model contained fewer groups than the Species Level model,
213 so that more training images were available for each group. We used both models because if the
214 Species Level model had poor accuracy, we predicted the Group Level model would have better
215 accuracy since more training images would be available for several groups. As it is the most
216 broadly applicable model and is the one implemented in the MLWIC package, we will mainly
217 discuss the Species Level model here, but show results from the Group Level to demonstrate
218 alternative approaches.

219 For each of the three models, 90% of the classified images for each species or group were used
220 to train the model and 10% of the images were used to test it in most cases. However, we wanted
221 to evaluate the model’s performance for each species present at each study site, so we altered
222 training-testing allocation for the rare situations where there were few classified images of a
223 species at a site. Specifically, with 1-9 classified images for a species at a site, we used all of
224 these images for testing and none for training; for site-species pairs with 10-30 images, 50%
225 were used for training and testing; and for > 30 images per site for each species, 90% were
226 allocated to training and 10% to testing (Appendices S3 - S7 show the number of training and
227 test images for each species at each site).

228

229 *Evaluating model accuracy*

230 Model testing was conducted by running the trained model on the withheld images that were not
231 used to train the model. Accuracy (A) was assessed as the proportion of images in the test dataset
232 (N) that were correctly classified (C) by the top guess ($A = C/N$). Top 5 accuracy ($A5$) was
233 defined as the proportion of images in the test dataset that were correctly classified by any of the
234 top 5 assignments ($C5$; $A5 = C5/N$). For each species or group we calculated the rate of false
235 positives (FP) as the proportion of images classified as this species or group ($N_{model\ group}$) by
236 the model's top guess that contained a different species according to human observers
237 ($N_{true\ other}$; $FP = N_{true\ other}/N_{model\ group}$). We calculated the rate of false negatives for each
238 species (FN) as the proportion of images observers classified as a specific species or group
239 ($N_{true\ group}$) that the model's top guess classified differently ($N_{model\ other}$; $FN =$
240 $N_{model\ other}/N_{true\ group}$). This assumes the observers were correct in their classification of
241 images. We fit generalized additive models (GAMs) to the relationship between accuracy and the
242 logarithm (base 10) of the number of images used to train the model. We also calculated the
243 accuracy and rates of error specific to each of the five data sets from which images were
244 acquired.

245 To evaluate how the model would perform for a completely new study site in North America, we
246 used a dataset of 5,900 classified images of ungulates (moose, cattle, elk, and wild pigs) from
247 Saskatchewan, Canada by running the Species Level model on these images. We also evaluated
248 the ability of the model to operate on images with a completely different species community
249 (from Tanzania) to determine the model's ability to correctly classify images as having an animal
250 or being empty when encountering new species that it has not been trained to recognize. This
251 was done using 3.2 million classified images from the Snapshot Serengeti dataset (Swanson et
252 al., 2015).

253

254 **Results**

255 Our models performed well, achieving $\geq 97.5\%$ accuracy of identifying the correct species with
256 the top guess (Table 2). The model determining presence or absence of wild pigs had the highest
257 accuracy of all of our models (98.6%; Pig/no pig; Table 2). For the Species Level and Group
258 Level models, the top 5 accuracy was $> 99.9\%$. The model confidence in the correct answer
259 varied, but was mostly $> 95\%$; see Fig. 2 for confidences for each image for three example
260 species. Supporting a similar finding for camera trap images in Norouzzadeh et al. (2018), and a
261 general trend in deep learning (Goodfellow et al., 2016), species and groups that had more
262 images available for training were classified more accurately (Fig. 3, Table 1). GAMs relating
263 the number of training images with accuracy predicted 95% accuracy could be achieved when
264 approximately 71,000 training images were available for a species or group. However, these
265 models were not perfect fits to the data, and for several species and groups, 95% accuracy was
266 achieved with fewer than 70,000 images (Fig. 3). We found there was not a large effect of
267 daytime vs. nighttime on accuracy in the Species Level model as daytime accuracy was 98.2%
268 and nighttime accuracy was 96.6%. The top 5 accuracies for both times of day were $\geq 99.9\%$.
269 When we subsetting the testing dataset by study site, we found that site-specific accuracies ranged
270 from 90-99% (Appendices S3 - S7). The model performed poorly (0 – 22% accuracy) for species
271 in the four instances when the model did not include training images from that site (when < 10
272 classified images were available for the species/study site combination; Appendices S3 - S7).
273 Upon further investigation, we found these images were difficult to classify manually. For
274 example, striped skunks in Florida were misclassified in both of the images from this study site

275 (Appendix S5). These images both contained the same individual at the same camera, and most
276 wildlife experts would not classify it as a skunk (Appendix S8).

277 When we conducted out-of-sample validation by using our model to evaluate images of
278 ungulates from Canada, we achieved an overall accuracy of 81.8% with a top 5 accuracy of
279 90.9%. When we tested the ability of our model to accurately predict presence or absence of an
280 animal in the image using the Serengeti Snapshot dataset, we found that 85.1% were classified
281 correctly as empty, while 94.3% of images containing an animal were classified as containing an
282 animal. Our trained model was capable of classifying approximately 2,000 images per minute on
283 a Macintosh laptop with 16 gigabytes (GB) of RAM.

284

285 **Discussion**

286 To our knowledge, our Species Level model achieved the highest accuracy (97.5%) to date in
287 using machine learning for wildlife image classification (a recent paper achieved 95% accuracy;
288 Norouzzadeh et al., 2018). This model performed almost as well during the night as during the
289 day (accuracy = 97% and 98%, respectively). We provide this model as an R package (MLWIC),
290 which is especially useful for researchers studying the species and groups available in this
291 package (Table 1) in North America, as it performed well in classifying ungulates in an out-of-
292 sample test of images from Canada. The model can also be valuable for researchers studying
293 other species by removing images without any animals from the dataset before beginning manual
294 classification, as we achieved high accuracy in separating empty images from those containing
295 animals in a dataset from Tanzania. This R package can also be a valuable tool for any

296 researchers that have classified images, as they can use the package to train their own model that
297 can then classify any subsequent images collected.

298

299 *Optimizing camera trap use and application in ecology*

300 The ability to rapidly identify millions of images from camera traps can fundamentally change
301 the way ecologists design and implement wildlife studies. Camera trap projects amass large
302 numbers of images which require a sizable time investment to manually classify. For example,
303 the Snapshot Serengeti project (Swanson et al., 2015) amassed millions of images and employed
304 28,000 volunteers to manually classify 1.5 million images (Swanson et al., 2016; Palmer et al.,
305 2017). We found researchers can classify approximately 200 images per hour. Therefore, a
306 project that amasses 1 million images would require 10,000 hours for each image to be doubly
307 observed. To reduce the number of images that need to be classified manually, ecologists using
308 camera traps often limit the number of photos taken by reducing the size of camera arrays,
309 reducing the duration of camera trap studies, and imposing limits on the number of photos a
310 camera takes (Kelly et al., 2008; Scott et al., 2018). This constraint can be problematic in many
311 studies, particularly those addressing rare or elusive species that are often the subject of
312 ecological studies (O'Connell et al., 2011), as these species often require more effort to detect
313 (Tobler et al., 2008). Using deep learning methods to automatically classify images essentially
314 eliminates one of the primary reasons camera trap arrays are limited in size or duration. The
315 Species Level model in our R package can accurately classify 1 million images in less than nine
316 hours with minimal human involvement.

317 Another reason to limit the number of photos taken by camera traps is storage limitations on
318 cameras (Rasambainarivo et al., 2017; Hanya et al., 2018). When classifying images manually,
319 we might try to use high resolution photos to improve technicians' abilities to accurately classify
320 images, but higher resolution photos require more storage on cameras. Our results show a model
321 can be accurately trained and applied using low-resolution (256 x 256 pixel) images, but many of
322 these images were re-sized from a higher resolution, which might contain more information than
323 those which originated at a low resolution. Nevertheless, we hypothesize a model can be
324 accurately trained using images from low resolution cameras, and our R package allows users
325 who have such images to test this hypothesis. If supported, this can make camera trap data
326 storage much more efficient. Typical cameras set for 2048 x 1536 pixel resolution will run out of
327 storage space when they reach approximately 1,250 photos per GB of storage. Taking low
328 resolution images instead can increase the number of photos stored per GB to about 10,000 and
329 thus decrease the frequency at which researchers must visit cameras to change storage cards by a
330 factor of eight. Minimizing human visitation also will reduce human scents and disturbances that
331 could deter some species from visiting cameras. In the future, it may be possible to implement a
332 machine learning model on a game camera (Elias et al., 2017) that automatically classifies
333 images as empty or containing animals so that empty images are discarded immediately and not
334 stored on the camera. This type of approach could dramatically reduce the frequency with which
335 technicians need to visit cameras. Furthermore, if models effectively use low-resolution images,
336 it is not necessary for researchers to purchase high resolution cameras. Instead, researchers can
337 purchase lower cost, lower resolution cameras and allocate funding toward purchasing more
338 cameras and creating larger camera arrays.

339

340 *Applications to management of invasive and sensitive species*

341 By removing some of the major burdens associated with the use of camera traps, our approach
342 can be utilized by ecologists and wildlife managers to conduct more extensive camera trapping
343 surveys than were previously possible. One potential use is in monitoring the distribution of
344 sensitive or invasive species. For example, the distribution of invasive wild pigs in North
345 America is commonly monitored using camera traps. Humans introduce this species into new
346 locations that are often geographically distant from their existing range (Tabak et al., 2017),
347 which can quickly lead to newly-established populations. Camera traps could be placed in areas
348 at risk for introduction and provide constant surveillance. An automated image classification
349 model that simply ‘looks’ for pigs in images could monitor camera trap images and alert
350 managers when images with pigs are found, facilitating removal of animals before populations
351 establish. Additionally, after wild pigs have been eradicated from a region, camera traps could be
352 used to monitor the area to verify eradication success and automatically detect re-colonization or
353 reintroduction events. Similar approaches can be used in other study systems to more rapidly
354 detect novel invasive species arrivals, track the effects of management interventions, monitor
355 species of conservation concern, or monitor sensitive species following reintroduction efforts.

356

357 *Limitations*

358 Using out-of-sample model validation on a dataset from Canada revealed a lower accuracy
359 (82%) than at study sites from which our model was trained. Additionally, when we did not
360 include images of species/site combinations in training the model, due to low sample sizes, the
361 model performed poorly (Appendices S3 - S7; but these images were often difficult to classify

362 even by wildlife experts, Appendix S8). One potential explanation is the model evaluated both
363 the animal and the environment in the image and these are confounded in the species
364 identification (Norouzzadeh et al., 2018). Therefore, the model may have lower accuracies in
365 environments that were not in the training dataset. Ideally, the training dataset would include
366 training images representing the range of environments in which a species exists. Our model
367 includes training images from diverse ecosystems, making it relevant for classifying images from
368 many locations in North America. A further limitation is in our reported overall accuracy, which
369 is reported across all of the images that were available for testing, and we had considerable
370 imbalance in the number of images per species (Table 1). We provide accuracies for each
371 species, so the reader can more directly inspect model accuracy. Finally, our model was trained
372 using images that were classified by human observers, which are capable of making errors
373 (O'Connell et al., 2011; Meek, Vernes, & Falzon, 2013), meaning some of the images in our
374 training dataset were likely misclassified. Supervised machine learning algorithms require such
375 training examples, and therefore we are unaware of a method for training such models without
376 the potential for human classification error. Instead, we must acknowledge that these models will
377 make mistakes due to imperfections in both human observation and model accuracy.

378

379 *Future directions*

380 As this new technology becomes more widely available, ecologists will need to decide how it
381 will be applied in ecological analyses. For example, when using machine learning model output
382 to design occupancy and abundance models, we can incorporate accuracy estimates that were
383 generated when conducting model testing. The error of a machine learning model in identifying a
384 species is similar to the problem of imperfect detection of wildlife when conducting field

385 surveys. Wildlife are often not detected when they are present (false negatives) and occasionally
386 detected when they are absent (false positives); ecologists have developed models to effectively
387 estimate occupancy when data have these types of errors (Royle & Link, 2006; Guillera-Arroita
388 et al., 2017). We can use Bayesian occupancy and abundance models where the central
389 tendencies of the prior distributions for the false negative and false positive error rates are
390 derived from testing the machine learning model (e.g., values in Table 1). While we would
391 expect false positive rates in occupancy models to resemble the false positive error rates for the
392 machine learning model, false negative error rates would be a function of the both the machine
393 learning model and the propensity for some species to avoid detection by cameras when they are
394 present (Tobler et al., 2015).

395 Another area in need of development is how to group taxa when few images are available for the
396 species. We grouped species when few images were available for model training using an
397 arbitrary cut off of approximately 2,000 images per group (Table 1). We had few images of
398 horses (*Equus* spp.), but the model identified these images relatively well (93% accuracy),
399 presumably because they are phenotypically different from other species in our dataset. We also
400 had few images of opossums (*Didelphis virginiana*), but we did not group this species because it
401 is phenotypically different from other species in our dataset and was of ecological interest in our
402 studies; we achieved lower accuracy for this species (78%). We also included a group for rodents
403 from species for which we only had few images (*Erethizon dorsatum*, *Marmota flaviventris*,
404 *Genomys* spp., *Mus* spp., *Neotoma* spp., *Peromyscus* spp., *Tamias* spp., and *Rattus* spp.). The
405 model achieved relatively low accuracy for this group (79%), presumably because there were
406 few images for training (3,279) and members of this group are phenotypically different, making
407 it difficult for the model to train on this group. When researchers develop new machine learning

408 models, they will need to consider the available data, the species or groups in their study, and the
409 ecological question that the model will help address.

410 Here, we mainly focused on the species or class that the model predicted with the highest
411 confidence (the top guess), but in many cases researchers may want to incorporate information
412 from the model's confidence in the guess and additional model guesses. For example, if we are
413 interested in the highest overall accuracy, we could only consider images where the confidence
414 in the top guess is $> 95\%$. If we subset the results from our model test in this manner, we remove
415 10% of the images, but total accuracy increases to 99.6%. However, if the objective of a project
416 is to identify rare species, researchers may want to focus on all images in which the model
417 predicts that species to be in the top 5 guesses (the 5 species or groups that the model predicts to
418 have the highest confidence). In our model test, the correct species was in the top 5 guesses in
419 99.9% of the images, indicating that this strategy may be viable.

420 We expect the performance of machine learning models to improve in the future (Jordan &
421 Mitchell, 2015), allowing ecologists to further exploit this technology. Our model required
422 manual identification of many images to obtain high levels of accuracy (Table 1). Our model was
423 also limited in that we were only able to classify the presence or absence of species; we were not
424 able to determine the number of individuals, their behavior, or demographics. Similar machine
425 learning models are capable of including the number of animals and their behavior in
426 classifications (Norouzzadeh et al., 2018), but we could not include these factors because they
427 were rarely recorded manually in our dataset. As machine learning techniques improve, we
428 expect models will require fewer manually classified images to achieve high accuracy in
429 identifying species, counting individuals, and specifying demographic information. Furthermore,
430 as scientists begin projects intending to use machine learning to classify images, they may be

431 more willing to spend time extracting detailed information from fewer images instead of
432 obtaining less information from all images. This development would create a larger dataset of
433 information from images that can be used to train models. As machine learning algorithms
434 improve and ecologists begin considering this technology when they design studies, we think
435 that many novel applications will arise.

436 As camera trap use is a common approach to studying wildlife worldwide, there are likely now
437 large datasets of classified images. If scientists work together and share these datasets, we can
438 create large image libraries that span continents (Steenweg et al., 2017); we may eventually be
439 able to train a machine learning model that can identify many global species and be used by
440 researchers globally. Further, effectively sharing images and classifications can potentially be
441 integrated with a web-based platform, similar to that employed by Camera Base
442 (<http://www.atrium-biodiversity.org/tools/camerabase>) or eMammal (<https://emammal.si.edu/>).

443

444 **Acknowledgements**

445 We thank the hundreds of volunteers and employees who manually classified images and
446 deployed camera traps. We thank Dan Walsh for facilitating cooperation amongst groups.
447 Camera trap projects were funded by the U.S. Department of Energy under award # DE-
448 EM0004391 to the University of Georgia Research Foundation; USDA Animal and Plant Health
449 Inspection Service, National Wildlife Research Center and Center for Epidemiology and Animal
450 Health; Colorado Parks and Wildlife; Canadian Natural Science and Engineering Research
451 Council; University of Saskatchewan; and Idaho Department of Game and Fish.

452

453 **Data Accessibility**

454 The trained Species Level model is available in the R package MLWIC from the github
455 repository mikeyEcology/MLWIC. Images used for training and testing models and their
456 classifications are available a digital repository.

457

458 **Author Contributions**

459 MAT, RSM, KCV, NPS, SJS, and DWW conceived of the project; DWW, JSL, MAT, RKB,
460 BW, PAD, JCB, MDW, BT, PES, NPS, KCV, JMH, ESN, JSI, EAO, RKB, PML, and AKM
461 oversaw collection and manual classification of wildlife in camera trap images from the study
462 sites; MSN and JC developed and programmed the machine learning model; MAT led the
463 analyses and writing of the R package; EGM assisted with R package development and
464 computing; MAT and RSM led the writing. All authors contributed critically to drafts and gave
465 final approval for submission.

466

467 **References**

468 Adabi, M., Barhab, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow:
469 a system for large-scale machine learning (Vol. 16, pp. 265–283). Presented at the 12th
470 USENIX Symposium on Operating Systems Design and Implementation, USENIX
471 Association.
472 Advanced Research Computing Center. (2012). *Mount Moran: IBM System X cluster*. Laramie,
473 WY: University of Wyoming. <https://arcc.uwyo.edu/guides/mount-moran>

- 474 Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., ... Kirschel, A.
475 N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays:
476 applications, technological considerations and prospectus: Acoustic monitoring. *Journal*
477 *of Applied Ecology*, 48(3), 758–767. doi:10.1111/j.1365-2664.2011.01993.x
- 478 Brown, J. H. (1995). *Macroecology*. University of Chicago Press.
- 479 Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014). Deep convolutional neural
480 network based species recognition for wild animal monitoring (pp. 858–862). IEEE
481 International Conference on Image Processing (ICIP). doi:10.1109/ICIP.2014.7025172
- 482 Elias, A. R., Golubovic, N., Krintz, C., & Wolski, R. (2017). Where’s the bear?: automating
483 wildlife image processing using IoT and Edge Cloud Systems (pp. 247–258). ACM
484 Press. doi:10.1145/3054977.3054986
- 485 Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species: The art
486 of modelling range-shifting species. *Methods in Ecology and Evolution*, 1(4), 330–342.
487 doi:10.1111/j.2041-210X.2010.00036.x
- 488 Getzin, S., Wiegand, K., & Schöning, I. (2012). Assessing biodiversity in forests using very
489 high-resolution images and unmanned aerial vehicles: Assessing biodiversity in forests.
490 *Methods in Ecology and Evolution*, 3(2), 397–404. doi:10.1111/j.2041-
491 210X.2011.00158.x
- 492 Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring:
493 Identification of animal species in camera-trap images using very deep convolutional
494 neural networks. *Ecological Informatics*, 41, 24–32. doi:10.1016/j.ecoinf.2017.07.004
- 495 Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (1st ed.). Cambridge,
496 Massachusetts: MIT Press.

- 497 Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R.
498 (2017). Dealing with false-positive and false-negative errors about species occurrence at
499 multiple levels. *Methods in Ecology and Evolution*, 8(9), 1081–1091. doi:10.1111/2041-
500 210X.12743
- 501 Gurney, K. (2014). *An Introduction to Neural Networks* (1st ed.). London: CRC Press. Retrieved
502 from <https://www.taylorfrancis.com/books/9781482286991>
- 503 Hanya, G., Otani, Y., Hongo, S., Honda, T., Okamura, H., & Higo, Y. (2018). Activity of wild
504 Japanese macaques in Yakushima revealed by camera trapping: Patterns with respect to
505 season, daily period and rainfall. *PLOS ONE*, 13(1), e0190631.
506 doi:10.1371/journal.pone.0190631
- 507 He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In
508 *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp.
509 770–778). IEEE. doi:10.1109/CVPR.2016.90
- 510 Howe, E. J., Buckland, S. T., Després-Einspenner, M.-L., & Kühl, H. S. (2017). Distance
511 sampling with camera traps. *Methods in Ecology and Evolution*, 8(11), 1558–1565.
512 doi:10.1111/2041-210X.12790
- 513 Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects.
514 *Science*, 349(6245), 255–260. doi:10.1126/science.aaa8415
- 515 Kelly, M. J., Noss, A. J., Di Bitetti, M. S., Maffei, L., Arispe, R. L., Paviolo, A., ... Di Blanco,
516 Y. E. (2008). Estimating Puma Densities from Camera Trapping across Three Study
517 Sites: Bolivia, Argentina, and Belize. *Journal of Mammalogy*, 89(2), 408–418.
518 doi:10.1644/06-MAMM-A-424R.1

- 519 Lazebnik, S., Schmid, C., & Ponce, J. (2006). Beyond bags of features: spatial pyramid matching
520 for recognizing natural scene categories. In *Computer vision and pattern recognition*
521 (Vol. 2, pp. 2169–2178). New York: IEEE.
- 522 MacArthur, R. H. (1984). *Geographical ecology: patterns in the distribution of species*.
523 Princeton, New Jersey: Princeton University Press.
- 524 MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E.
525 (2017). *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of*
526 *Species Occurrence* (2nd ed.). London, UK: Academic Press.
- 527 Meek, P. D., Vernes, K., & Falzon, G. (2013). On the reliability of expert identification of small-
528 medium sized mammals from camera trap photos. *Wildlife Biology in Practice*, 9(2).
529 doi:10.2461/wbp.2013.9.4
- 530 Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2012). *Foundations of Machine Learning*. MIT
531 Press.
- 532 Niedballa, J., Sollmann, R., Courtiol, A., & Wilting, A. (2016). camtrapR : an R package for
533 efficient camera trap data management. *Methods in Ecology and Evolution*, 7(12), 1457–
534 1462. doi:10.1111/2041-210X.12600
- 535 Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., &
536 Clune, J. (2018). Automatically identifying, counting, and describing wild animals in
537 camera-trap images with deep learning. *Proceedings of the National Academy of*
538 *Sciences*, 201719367. doi:10.1073/pnas.1719367115
- 539 O’Connell, A. F., Nichols, J. D., & Karanth, K. U. (Eds.). (2011). *Camera traps in animal*
540 *ecology: methods and analyses*. Tokyo ; New York: Springer.

- 541 Palmer, M. S., Fieberg, J., Swanson, A., Kosmala, M., & Packer, C. (2017). A ‘dynamic’
542 landscape of fear: prey responses to spatiotemporal variations in predation risk across the
543 lunar cycle. *Ecology Letters*, 20(11), 1364–1373. doi:10.1111/ele.12832
- 544 Rasambainarivo, F., Farris, Z. J., Andrianalizah, H., & Parker, P. G. (2017). Interactions between
545 carnivores in Madagascar and the risk of disease transmission. *EcoHealth*, 14(4), 691–
546 703. doi:10.1007/s10393-017-1280-7
- 547 Rovero, F., Zimmermann, F., Bersi, D., & Meek, P. (2013). ‘Which camera trap type and how
548 many do I need?’ A review of camera features and study designs for a range of wildlife
549 research applications. *Hystrix, the Italian Journal of Mammalogy*, 24(2), 1–9.
- 550 Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology*. New
551 York: Academic Press.
- 552 Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false
553 positive and false negative errors. *Ecology*, 87(4), 835–841.
- 554 Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-
555 propagating errors. *Nature*, 323(6088), 533–536. doi:10.1038/323533a0
- 556 Scott, A. B., Phalen, D., Hernandez-Jover, M., Singh, M., Groves, P., & Toribio, J.-A. L. M. L.
557 (2018). Wildlife presence and interactions with chickens on Australian commercial
558 chicken farms assessed by camera traps. *Avian Diseases*, 62(1), 65–72.
559 doi:10.1637/11761-101917-Reg.1
- 560 Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image
561 recognition. *ArXiv:1409.1556 [Cs]*. Retrieved from <http://arxiv.org/abs/1409.1556>
- 562 Steenweg, R., Hebblewhite, M., Kays, R., Ahumada, J., Fisher, J. T., Burton, C., ... Rich, L. N.
563 (2017). Scaling-up camera traps: monitoring the planet’s biodiversity with networks of

- 564 remote sensors. *Frontiers in Ecology and the Environment*, 15(1), 26–34.
- 565 doi:10.1002/fee.1448
- 566 Swanson, A., Kosmala, M., Lintott, C., & Packer, C. (2016). A generalized approach for
567 producing, quantifying, and validating citizen science data from wildlife images.
568 *Conservation Biology*, 30(3), 520–531. doi:10.1111/cobi.12695
- 569 Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot
570 Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an
571 African savanna. *Scientific Data*, 2, 150026. doi:10.1038/sdata.2015.26
- 572 Swinnen, K. R. R., Reijniers, J., Breno, M., & Leirs, H. (2014). A novel method to reduce time
573 investment when processing videos from camera trap studies. *PLoS ONE*, 9(6), e98881.
574 doi:10.1371/journal.pone.0098881
- 575 Tabak, M. A., Piaggio, A. J., Miller, R. S., Sweitzer, R. A., & Ernest, H. B. (2017).
576 Anthropogenic factors predict movement of an invasive species. *Ecosphere*, 8(6),
577 e01844. doi:10.1002/ecs2.1844
- 578 Tikhonov, G., Abrego, N., Dunson, D., & Ovaskainen, O. (2017). Using joint species
579 distribution models for evaluating how species-to-species associations depend on the
580 environmental context. *Methods in Ecology and Evolution*, 8(4), 443–452.
581 doi:10.1111/2041-210X.12723
- 582 Tobler, M. W., Carrillo-Percestequi, S. E., Leite Pitman, R., Mares, R., & Powell, G. (2008). An
583 evaluation of camera traps for inventorying large- and medium-sized terrestrial rainforest
584 mammals. *Animal Conservation*, 11(3), 169–178. doi:10.1111/j.1469-1795.2008.00169.x

- 585 Tobler, M. W., Zúñiga Hartley, A., Carrillo-Percastegui, S. E., & Powell, G. V. N. (2015).
586 Spatiotemporal hierarchical modelling of species richness and occupancy using camera
587 trap data. *Journal of Applied Ecology*, 52(2), 413–421. doi:10.1111/1365-2664.12399
- 588 Underwood, A. ., Chapman, M. ., & Connell, S. . (2000). Observations in ecology: you can't
589 make progress on processes without understanding the patterns. *Journal of Experimental*
590 *Marine Biology and Ecology*, 250(1–2), 97–115. doi:10.1016/S0022-0981(00)00181-7
- 591 Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. (2013). Automated
592 identification of animal species in camera trap images. *EURASIP Journal on Image and*
593 *Video Processing*, 2013(1). doi:10.1186/1687-5281-2013-52
- 594
- 595

Tables and Figures

Table 1: Accuracy of the Species Level model

Species or group name	Scientific name	Number of training images	Number of test images	Accuracy	Top 5 accuracy	False positive rate	False negative rate
Moose	<i>Alces alces</i>	8,967	997	0.98	1.00	0.02	0.02
Cattle	<i>Bos taurus</i>	1,817,109	201,903	0.99	1.00	0.01	0.01
Quail	<i>Callipepla californica</i>	2,039	236	0.90	0.96	0.11	0.10
Canidae	Canidae	20,851	2,321	0.89	0.99	0.08	0.11
Elk	<i>Cervus canadensis</i>	185,390	20,606	0.98	1.00	0.01	0.02
Mustelidae	Mustelidae	1,991	223	0.76	0.98	0.12	0.24
Corvid	Corvidae	4,037	452	0.79	1.00	0.15	0.21
Armadillo	<i>Dasypus novemcinctus</i>	8,926	993	0.87	0.99	0.08	0.13
Turkey	<i>Meleagris gallopavo</i>	3,919	447	0.88	1.00	0.12	0.12
Opossum	<i>Didelphis virginiana</i>	1,804	210	0.78	0.96	0.15	0.22
Horse	<i>Equus</i> spp.	2,517	281	0.93	0.99	0.05	0.07
Human	<i>Homo sapiens</i>	88,667	9,854	0.96	1.00	0.03	0.04
Rabbits	Leporidae	17,768	1,977	0.96	1.00	0.06	0.04
Bobcat	<i>Lynx rufus</i>	22,889	2,554	0.90	0.99	0.05	0.10
Striped skunk	<i>Mephitis mephitis</i>	10,331	1,154	0.95	0.99	0.03	0.05
Unidentified deer	<i>Odocoileus</i> spp.	86,502	9,613	0.96	1.00	0.02	0.04
Rodent	Rodentia	3,279	366	0.79	0.98	0.17	0.21
Mule deer	<i>Odocoileus hemionus</i>	76,878	8,543	0.98	1.00	0.03	0.02
White-tailed deer	<i>Odocoileus virginianus</i>	12,238	1,360	0.81	1.00	0.22	0.19
Raccoon	<i>Procyon lotor</i>	42,948	4,781	0.88	1.00	0.10	0.12

Mountain lion	<i>Puma concolor</i>	13,272	1,484	0.93	0.98	0.03	0.07
Squirrel	<i>Sciurus spp.</i>	59,072	6,566	0.96	1.00	0.05	0.04
Wild pig	<i>Sus scrofa</i>	287,017	31,893	0.97	1.00	0.02	0.03
	<i>Vulpes vulpes</i> and <i>Urocyon</i>						
Fox	<i>Cinereoargenteus</i>	10,749	1,204	0.91	0.99	0.07	0.09
Black Bear	<i>Ursus americanus</i>	79,628	8,850	0.94	1.00	0.02	0.06
Vehicle		23,413	2,602	0.93	1.00	0.04	0.07
Bird	Aves	61,063	6,787	0.94	1.00	0.05	0.06
Empty		414,119	46,016	0.96	1.00	0.06	0.04
Total		3,367,383	374,273	0.98	1.00		

Table 2: Accuracy (across all images for all species) of the three deep learning tasks analyzed

Model	Accuracy (%)
Pig/no pig	98.6
Species Level	97.5
Group Level	97.8

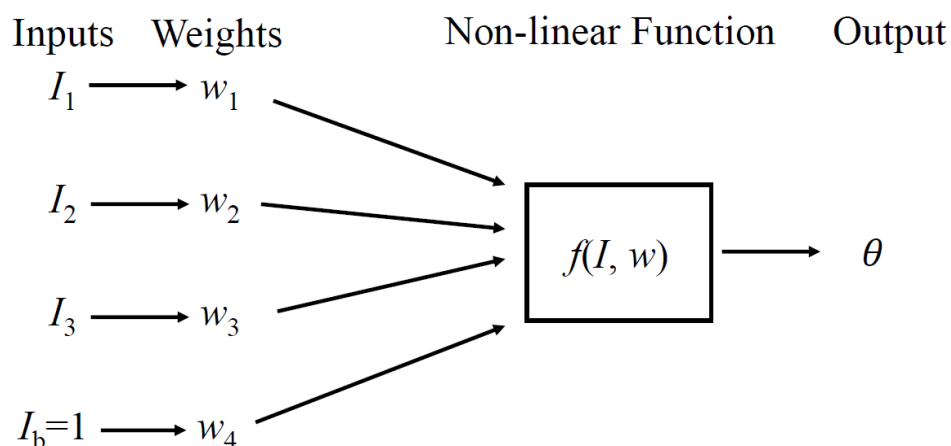


Figure 1: Within an artificial neural network, inputs (I) are multiplied by their weights (w), summed, and then evaluated by a non-linear function, which also accounts for bias (I_b). The output (θ) can be passed as input into other neurons or serve as network outputs.

Backpropagation involves adjusting the weights so that a model can provide the desired output.

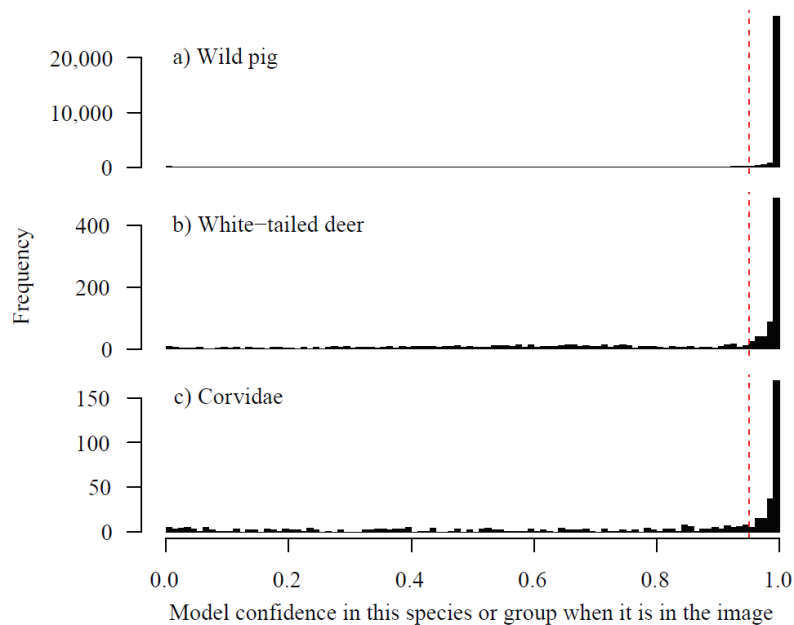


Fig. 2: Histograms represent the confidence assigned by all of the top five guesses by the Species Level model for each of these three example species when it was present in an image. The dashed line represents 95% confidence; the majority of model-assigned confidences were greater than this value.

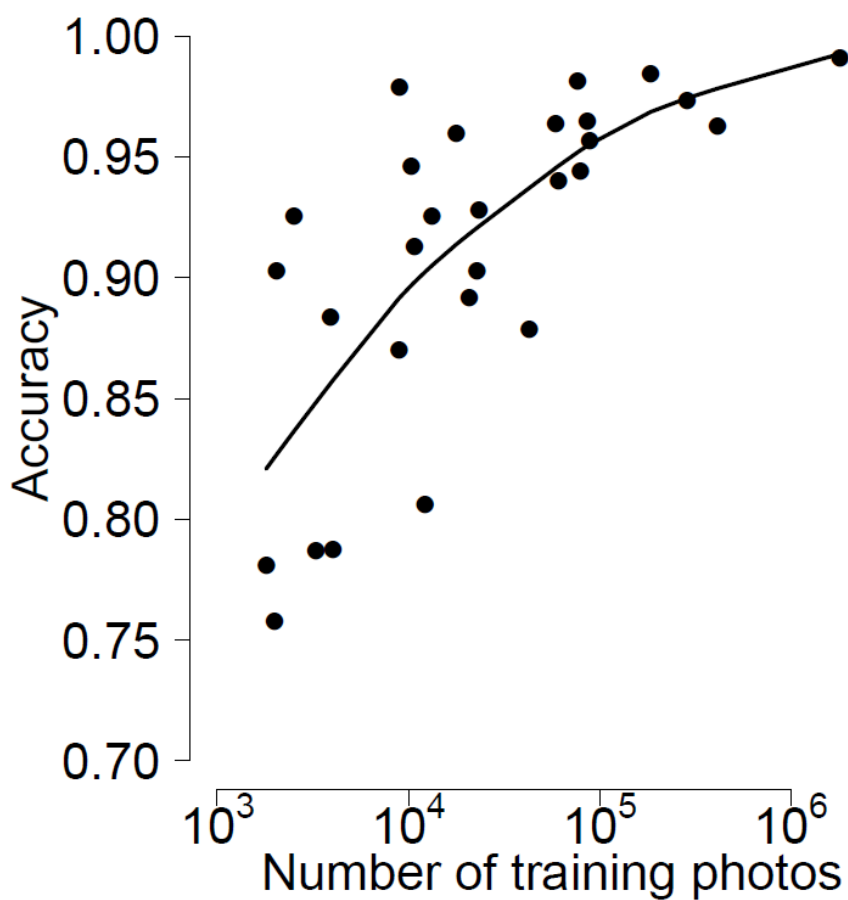


Fig. 3: Machine learning model accuracy increased with the size of the training dataset. Points represent each species or group of species. The line represents the result of generalized additive models relating the two variables.

Supporting Information

Appendix S1. Site descriptions for each of the study locations

Appendix S2. Accuracy of the Group Level for each species

Appendix S3. Accuracy of the Species Level model at the Tejon research site in California.

Appendix S4. Accuracy of the Species Level model in Colorado

Appendix S5. Accuracy of the Species Level model at Buck Island Ranch in Florida

Appendix S6. Accuracy of the Species Level model at the Camp Bullis Military Training Center
in Texas

Appendix S7. Accuracy of the Species Level model at the Savannah River Ecology Laboratory
in South Carolina

Appendix S8. Example of a striped skunk that was misclassified